Data Mining Using Rules Extracted from SVM: An Application to Churn Prediction in Bank Credit Cards

M.A.H. Farquad[1,2], V. Ravi[1,*] , and S. Bapi Raju[2]

[1] Institute for Development and Research in Banking Technology, Castle Hills Road #1, Masab Tank, Hyderabad – 500 057 (AP) India
[2] Department of Computer & Information Sciences, University of Hyderabad, Hyderabad – 500 046 (AP) India
Tel.: +91-40-2353 4981
farquadonline@gmail.com, rav_padma@yahoo.com, bapics@uohyd.ernet.in

Abstract. In this work, an eclectic procedure for rule extraction from Support Vector Machine is proposed, where Tree is generated using Naïve Bayes Tree (NBTree) resulting in the SVM+NBTree hybrid. The data set analyzed in this paper is about churn prediction in bank credit cards and is obtained from Business Intelligence Cup 2004. The data set under consideration is highly unbalanced with 93.11% loyal and 6.89% churned customers. Since identifying churner is of paramount importance from business perspective, sensitivity of classification model is more critical. Using the available, original unbalanced data only, we observed that the proposed hybrid SVM+NBTree yielded the best sensitivity compared to other classifiers.

Keywords: Churn prediction in credit cards, Support Vector Machine, Rule Extraction, Naive Bayes Tree.

1 Introduction

Data mining (also called as Knowledge Discovery in Database) is a process that consists of applying data analysis and discovery algorithms that produce particular enumeration of pattern (or model) over the data [1]. Data mining has been efficiently used in wide range of profiling practices, such as manufacturing [2], fraud detection [3].

Increasing number of customers has made the banks conscious of the quality of the services they offer. The problem of customers shifting loyalties from one bank to another has become common. This phenomenon, called '*churn*' occurs due to reasons such as availability of latest technology, customer-friendly staff and proximity of geographical location, etc. Hence, there is a pressing need to develop models that can predict which existing '*loyal*' customer is going to churn out in near future [4].

Research shows that customers with longer relationships with the firm have higher prior cumulative satisfaction ratings [5] than online bank customers [6]. It is more profitable to segment and target customers on the basis of their (changing) purchase

* Corresponding author.

H. Sakai et al. (Eds.): RSFDGrC 2009, LNAI 5908, pp. 390–397, 2009.

behavior and service experiences rather than on the basis of their (stable) demographics [7]. Churn management consists of developing techniques that enable firms to keep their profitable customers and aims at increasing customer loyalty [8]. Churn prediction and management is one of the important activities of Customer Relationship Management.

Credit Card Database and PBX Database with Data Mining by Evolutionary Learning (DMEL) [9], emphasize on the natural differences between Savings and Investment (SI) products in banks to cross-sell in terms of both maximizing the customers' retention proneness and their preferences [10]. Chu et al., [11] reported that management should prepare an anti-churn strategy that is usually far less expensive than acquiring new customers.

In this paper we address an important issue of rule extraction from SVM and investigate its usefulness in credit card churn prediction in banks. The proposed approach is carried out in two major steps. (1) Extraction of support vectors and obtaining predictions for training instances and support vectros. (2) Rule generation. Incidentally, by using the predictions of SVM for training set and support vectors, we ensure that the rules generated are basically mimicking the behavior of SVM.

The rest of the paper is organized as follows. Section 2 presents the literature review of rule extraction from SVM approaches. Section 3 describes the data set used. Section 4 presents the proposed hybrid rule extraction approach. Section 5 presents results and discussion. Finally section 6 concludes the paper.

2 Rule Extraction from SVM

SVM [13] recently became one of the most popular classification methods. They have been used in wide variety of applications such as Text classification [14], Facial expression recognition [15], and Gene analysis [16] so on. Despite superior performance of SVM, they are often regarded as black box models. Converting this black box, high accurate models to transparent model is "*Rule Extraction*" [17].

Recently attempts have been made to extract rules from SVMs [18]. Intensive work has been done towards developing rule extraction techniques for neural networks but less work has been done for extracting rules from SVM. Some of the approaches proposed towards rule extraction from SVM are; SVM+Prototype [19], RulExtSVM [20], Extracting rules from trained support vector machines [21], Hyper rectangle Rules Extraction (HRE) [22], Fuzzy Rule Extraction (FREx) [23], Multiple Kernel-Support Vector Machine (MK-SVM) [24], SQRex-SVM [25], sequential covering approach [26] and Recently a new Active Learning-Based Approach (ALBA) [27] are some of the approaches proposed towards rule extraction from SVM.

Incidentally Farquad et al. [28, 29] also proposed a hybrid rule extraction approach using SVM and the extracted rules are tested for bankruptcy prediction in banks. They first extracted the support vectors then they used these support vectors to train Fuzzy Rule Based System (FRBS), Decision Tree and Radial Basis Function Network. They concluded that the hybrid SVM+FRBS outperformed the stand-alone classifiers.

3 Data Set Description

The dataset is from a Latin American bank that suffered from an increasing number of churns with respect to their credit card customers and decided to improve its retention system. Two groups of variables are available for each customer: sociodemographic and behavioural data, which are described in Table 1. The dataset comprises 22 variables, with 21 predictor variables and 1 class variable. It consists of 14814 records, of which 13812 are nonchurners and 1002 are churners, which means there are 93.24% nonchurners and 6.76% churners. Hence, the dataset is highly unbalanced in terms of the proportion of churners versus non-churners [30].

Table 1. Feature description of churn prediction data set

Feature	Description	Value
Target	Target Variable	0-NonChurner 1-Churner
CRED_T	Credit in month T	Positive real number
CRED_T-1	Credit in month T-1	Positive real number
CRED_T-2	Credit in month T-2	Positive real number
NCC_T	Number of credit cards in months T	Positive integer value
NCC_T-1	Number of credit cards in months T-1	Positive integer value
NCC_T-2	Number of credit cards in months T-2	Positive integer value
INCOME	Customer's Income	Positive real number
N_EDUC	Customer's educational level	1 - University student 2 - Medium degree 3 - Technical degree 4 - University degree
AGE	Customer's age	Positive integer
SX	Customers sex	1 - male 0 - Female
E_CIV	Civilian status	1-Single 2-Married 3-Widow 4-Divorced
T_WEB_T	Number of web transaction in months T	Positive integer
T_WEB_T-1	Number of web transaction in months T-1	Positive integer
T_WEB_T-2	Number of web transaction in months T-2	Positive integer
MAR_T	Customer's margin for the company in months T	Real Number
MAR_T-1	Customer's margin for the company in months T-1	Real Number
MAR_T-2	Customer's margin for the company in months T-2	Real Number
MAR_T-3	Customer's margin for the company in months T-3	Real Number
MAR_T-4	Customer's margin for the company in months T-4	Real Number
MAR_T-5	Customer's margin for the company in months T-5	Real Number
MAR_T-6	Customer's margin for the company in months T-6	Real Number

4 Proposed Rule Extraction Approach

In this research work we propose a hybrid rule extraction procedure for solving large scale classification problem in the framework of data mining using rules extracted from support vector machine. In the churn prediction problem, sensitivity alone is the important criteria, higher the sensitivity better is the model. The proposed hybrid is composed of two major steps (i) support vector extraction and obtaining the predictions of the training instances and support vectors extracted (ii) rule generation using

NBTree [12]. The data set used in this study is highly unbalanced. However, we did not employ any balancing technique to balance the data.

The current approach in this paper is distinct from earlier studies where this data set was analyzed [4, 31] in the following ways:

- Extraction of the support vectors makes the sample size very much small.
- Application of the approach is extended to a data mining problem.

Also, the hybrid approach presented here is different from [28, 29] in the following ways:

- Dealing with unbalanced large scale data set.
- Using the predictions of support vectors using SVM model i.e. *Case-SP* to generate rules with NBTree.

4.1 Support Vectors Extraction and Predictions of SVM

Figure 1 depicts the extraction of support vectors and the resulting 3 variants of the hybrid. The predictions for training set and support vectors are obtained from the developed SVM model. Here *Case-A* and *Case-SA* are two sets viz. Training and SVs set with their corresponding actual target values respectively. Predictions of SVM are obtained for *Case-A* and *Case-SA* and the actual target values are then replaced by the predicted target values to get *Case-P* and *Case-SP* respectively. By using the newly generated *Case-P* and *Case-SP* we ensure that the rules extracted are actually from SVM.

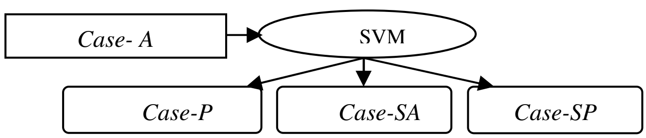

Fig. 1. Support vectors extraction and predictions of SVM

4.2 Rule Generation

During rule generation phase, depicted in Fig. 2, we analyzed 4 different data sets, (i) *Case-A*, (ii) *Case-P*, (iii) *Case-SA* and (iv) *Case-SP* separately. Rules are generated using NBtree hybrid [12]. NBTree attempts to utilize the advantages of both decision trees (i.e. segmentation) and naïve bayes (evidence accumulation from multiple attributes). A decision tree is built with univariate splits at each node, but with Navie-Bayes classifiers at the leaves. Rules are generated under 10-fold cross validation method and the average sensitivity is presented.

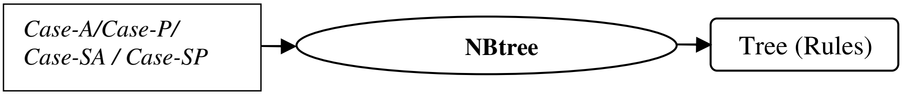

Fig. 2. Rule generation phase

4.3 Experimental Setup

The available large scale unbalanced dataset is first divided into two parts of 70:30 ratios. 70% of the data is then used for 10-Fold Cross Validation (10-FCV) and 30% of the data is named as validation set and stored for evaluating the efficiency of the rules generated using 10-FCV at later point of time. The class distribution in the training and validation data sets is as same as that in the original data i.e. 93.11% for loyal customers and 6.89% for churned customers. The accuracy and validity of the rules generated during 10-FCV are then tested against the validation set.

5 Results and Discussions

We used the SVM library viz., LibSVM [32] for building SVM model and support vector extraction. LibSVM is integrated software for support vector classification and is developed in MATLAB. RapidMiner4.5 community edition [33] is used for generating NBTree. Many business decision makers, dealing with churn prediction problem, place high emphasis on sensitivity alone because higher sensitivity leads to greater success in identifying potential churners correctly and thereby contributing to the bottom-line of the CRM viz., retaining extant loyal customers. Consequently in this paper, sensitivity is accorded top priority ahead of specificity. The results of hybrid in various cases (*Case-A, Case-P, Case-SA and Case-SP*) are presented in Table. 2.

Table 2. Average results of 10-fold cross validation

Classifier	*Test under 10-FCV*			*Validation*		
	Sens*	Spec*	Acc*	Sens*	Spec*	Acc*
SVM (*Case-A*)	64.2	74.86	74.13	60.17	74.92	73.92
NBTree (*Case-A*)	55.5	98.99	96.06	61.21	99.02	96.46
SVM + NBTree (*Case-P*)	*68.62*	78.45	77.78	*68.52*	78	77.07
SVM + NBTree (*Case-SA*)	0	100	93.11	0	100	93.11
SVM + NBTree (*Case-SP*)	*68.33*	74.38	75.18	*68.04*	75.34	75.11
Kumar and Ravi(2008) [9]	62.07	98.51	96.05	NA	NA	NA
Naveen et al. (2009) [10]	41.62	79.6	77.03	NA	NA	NA

Note: * Sens = sensitivity; Spec = specificity; Acc = accuracy

It is observed from the results that the hybrid SVM+NBTree using *Case-P* yielded the average sensitivity under 10-FCV and against validation set is 68.62% and 68.52% respectively. The hybrid SVM+NBTree using *Case-SP* obtained the average sensitivity under 10-FCV and against validation set is 68.33% and 68.04% respectively. Stand alone SVM and NBTree using *Case-A* yielded the average sensitivity of 64.2% and 55.5% respectively.

Working on the same data set, Kumar and Ravi [5] reported 62.07% average sensitivity achieved using decision tree classifier, whereas, Naveen et al. [31] reported 41.62% average sensitivity obtained using FuzzyARTMAP. Our results are not strictly comparable to their results as they performed 10-FCV without partitioning the

Table 3. Rule set extracted by SVM+NBTree hybrid using *Case-SP*

#	*Rule Antecedents*	*Consequent*
1	CRED_T<=598.1 and MAR_T-2<=14.045	Churner
2	CRED_T<=598.1 and MAR_T-2<=14.045 and INCOME<=1035 and MAR_T-4<=15.135 and T_WEB-T<=7.5	Churner
3	CRED_T>598.1 and T_WEB_T-2<=2.5	Non-Churner
4	CRED_T<=598.1 and MAR_T-2<=14.045 and E_CIV>1.5 and NCC_T-2>0.5 and T_WEB-T>7.5	Non-Churner
5	CRED_T<=598.1 and MAR_T-2<=14.045 and E_CIV>1.5 and NCC_T-2>0.5	Non-Churner
6	CRED_T<=598.1 and MAR_T-2<=14.045 and MAR_T-5<=9.73	Churner
7	CRED_T<=598.1 and MAR_T-2<=14.045 and E_CIV>1.5 and NCC_T-2>0.5 and INCOME>1035 and N_EDUC<=3.5	Non-Churner

original data set into training and validation set. From the above discussions, it is observed that the proposed hybrid SVM+NBTree using *Case-P* and *Case-SP* stand as the best performers compared to other classifiers evaluated in this study. The example rule set obtained by the hybrid SVM+NBTree using *Case-SP* is presented in Table 3.

The tree obtained using NBTree has naïve bayes classifiers at leaf nodes that indicates the probability of each class available in the data set used, instead of prediction of any single class. For better understanding of the tree we modified the rules and the class with higher probability assigned by the naïve bayes classifier at leaf node is considered the consequent of the rule. The number of rules extracted using our approach i.e. SVM+NBTree is very much less and rule length is smaller when compared to those of Kumar and Ravi [4].

It is observed that the number of SVs extracted is 67.8% less than the actual number of training instances. Still, we got a decent sensitivity of 68.52% in *Case-SP* which is a significant outcome of the present study. Hence it is recommended to use support vectors instead of using all the training instances to generate rules.

6 Conclusions

In this paper, we present a rule extraction approach from SVM using NBTree to solve customer churn prediction problem concerning bank credit cards. The data set is highly unbalanced data with 93.11% loyal customers and 6.89% churned customers. We did not employ any balancing technique for balancing the data. Instead we analyzed the original data. We infer that the proposed approach SVM+NBTree using *Case-P* and *Case-SP* outperformed all other classifiers tested and achieved best average sensitivity of 68.52% and 68.03% respectively. The following recommendations are offered from this work. (i) it is better to extract support vectors and use *Case-SP* to generate rules because the number of instances is drastically reduced in the form of support vectors, (ii) the resultant sensitivity yielded by *Case-SP* is almost similar to that of the sensitivity yielded by *Case-P*, (iii) time taken for generation of rules is cut short by more than 60% and (iv) the number of rules extracted and the antecedents per rule are small thereby improving the comprehensibility of the rules.

References

1. Usama, F., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery in Databases. American Association for Artificial Intelligence (1996)
2. Ravi, V., Arul Shalom, S.A., Manickavel, A.: Sputter Process Variables Prediction via Data Mining. In: Proceedings of the 2004 IEEE Conference on Cybernetics and Intelligent Systems, Singapore (2004)
3. Senator, T., Goldberg, H.G., Wooton, J., Cottini, M.A., Umarkhan, A.F., Klinger, C.D., Llamas, W.M., Marrone, M.P., Wong, R.W.H.: The Financial Crimes Enforcement Network AI System (FAIS): Identifying Potential Money Laundering from Reports of Large Cash Transactions. AI Magazine 16(4), 21–39 (1995)
4. Kumar, D.A., Ravi, V.: Predicting credit card customer churn in banks using data mining. International Journal for Data Analysis, Techniques and Strategies 1(1), 4–28 (2008)
5. Bolton, R.N.: A Dynamic model of the Duration of the customer's relationship with a continuous service provider: The Role of Satisfaction. Marketing Science 17(1), 45–65 (1998)
6. Mols, N.P.: The Behavioral consequences of PC banking. International Journal of Bank Marketing 16(5), 195–201 (1998)
7. Bolton, R.N., Kannan, P.K., Bramlett, M.D.: Implications of Loyalty Program Membership and Service Experiences for Customer Retention and Value. Journal of the Academy of Marketing Science 28(1), 95–108 (2000)
8. Lejeune, M.A.P.M.: Measuring the impact of data mining on churn management. Electronic Networking Applications and Policy 11(5), 375–387 (2001)
9. Au, W.-H., Chan, K.C.C., Yao, S.: A Novel Evolutionary Data Mining Algorithm with Applications to Churn Prediction. IEEE Transactions on Evolutionary Computation 7(6), 532–545 (2003)
10. Larivie're, B., Van den Poel, D.: Predicting customer retention and profitability by using random forests and regression forests techniques. Expert Systems with Applications 29(2), 472–484 (2005)
11. Chu, B.H., Tsai, M.-S., Ho, C.-S.: Toward a hybrid data mining model for customer retention. Knowledge-Based Systems 20(8), 703–718 (2007)
12. Ron, K.: Scaling Up the Accuracy of Naïve-Bayes Classifiers: a Decisoin-Tree Hybrid. In: Proceedings of KDD 1996, Portland, USA (1996)
13. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995)
14. Joachims, T.: Text categorization with support vector machines: Learning with many relevant features. In: Proceedings of the European Conference on Machine Learning. Springer, Heidelberg (1998)
15. Michel, P., Kaliouby, R.E.: Real time facial expression recognition in video using support vector machines. In: Proceedings of ICMI 2003, Vancouver, British Columbia, Canada, November 5-7, pp. 258–264 (2003)
16. Guyon, I., Weston, J., Barnhill, S., Vapnik, V.N.: Gene selection for cancer classification using support vector machines. Machine Learning 46(1-3), 389–422 (2002)
17. Gallant, S.: Connectionist expert systems. Communications of the ACM 31(2), 152–169 (1988)
18. Barakat, N.H., Diederich, J.: Eclectic rule-extraction from support vector machines. International journal of Computer Intelligence 2(1), 59–62 (2005)
19. Nunez, H., Angulo, C., Catata, A.: Rule extraction from support vector machines. In: European Symposium on Artificial Neural Networks Proceedings, pp. 107–112 (2002)

20. Fung, G., Sandilya, S., Bharat, R.R.: Rule extraction from linear support vector machines. In: Proceeding of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, pp. 32–40. ACM Press, New York (2005)
21. Barakat, N.H., Diederich, J.: Learning-based Rule-Extraction from Support Vector Machines. In: Proceedings of the 14th International Conference on Computer Theory and applications ICCTA 2004, Alexandria, Egypt (2004)
22. Zhang, Y., Su, H., Jia, T., Chu, J.: Rule Extraction from Trained Support Vector Machines. In: Ho, T.-B., Cheung, D., Liu, H. (eds.) PAKDD 2005. LNCS (LNAI), vol. 3518, pp. 61–70. Springer, Heidelberg (2005)
23. Barakat, N.H., Bradley, A.P.: Rule Extraction from Support Vector Machines: Measuring the Explanation Capability Using the Area under the ROC Curve. In: The 18th International Conference on Pattern Recognition (ICPR 2006), Hong Kong (2006)
24. Chen, Z., Li, J., Wei, L.: A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue. Artificial Intelligence in Medicine 41, 161–175 (2007)
25. Chaves, Ad.C.F., Vellasco, M.M.B.R., Tanscheit, R.: Fuzzy rule extraction from support vector machines. In: Fifth International Conference on Hybrid Intelligent Systems, Rio de Janeiro, Brazil, November 06-09 (2005)
26. Barakat, N.H., Bradley, A.P.: Rule Extraction from Support Vector Machines: A Sequential Covering Approach. IEEE Transactions on Knowledge and Data Engineering 19(6), 729–741 (2007)
27. Martens, D., Baesens, B., Gestel, T.V.: Decompositional Rule Extraction from Support Vector Machines by Active Learning. IEEE Transactions on Knowledge and Data Engineering 21(2), 178–191 (2009)
28. Farquad, M.A.H., Ravi, V., Bapi, R.S.: Support Vector Machine based Hybrid Classifiers and Rule Extraction Thereof: Application to Bankruptcy Prediction in Banks. In: Soria, E., Martín, J.D., Magdalena, R., Martínez, M., Serrano, A.J. (eds.) Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques. IGI Global (2008)
29. Farquad, M.A.H., Ravi, V., Bapi, R.S.: Rule Extraction using Support Vector Machine Based Hybrid Classifier. In: Presented in TENCON-2008, IEEE region 10 Conference, Hyderabad, India, November 19-21 (2008)
30. Business Intelligence Cup-2004: Organized by the Univeristy of Chile, `http://www.tis.cl/bicup_04/text-bicup/BICUP/202004/20public/20data.zip`
31. Naveen, N., Ravi, V., Kumar, D.A.: Application of fuzzyARTMAP for churn prediction in bank credit cards. International Journal of Information and Decision Sciences 1(4), 428–444 (2009)
32. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001) Software, `http://www.csie.ntu.edu.tw/~cjlin/libsvm`
33. Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., Euler, T.: YALE: Rapid Prototyping for Complex Data Mining Tasks. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2006 (2006)