Analytical CRM in banking and finance using SVM: a modified active learning-based rule extraction approach

M.A.H. Farquad

Institute for Development and Research in Banking Technology,
Castle Hills Road #1, Masab Tank,
Hyderabad – 500 057 (AP), India
and
Department of Computer and Information Sciences,
University of Hyderabad,
Hyderabad – 500 046 (AP), India
and
School of Business,
The University of Hong Kong,
Room 729(3), Meng Wah Complex,
Pokfulam Road, Hong Kong
E-mail: farquadonline@gmail.com

V. Ravi*

Institute for Development and Research in Banking Technology,
Castle Hills Road #1, Masab Tank,
Hyderabad – 500 057 (AP), India
E-mail: rav_padma@yahoo.com
*Corresponding author

S. Bapi Raju

Department of Computer and Information Sciences,
University of Hyderabad,
Hyderabad – 500 046 (AP), India
E-mail: bapics@uohyd.ernet.in

Abstract: This paper presents advancement to modified active learning-based approach in an eclectic framework for extracting *if-then* rules from support vector machine (SVM) for customer relationship management (CRM) purposes. The proposed approach comprises of three major phases:

1 *feature selection* using SVM-RFE (recursive feature elimination)
2 *active learning* for synthetic data generation
3 *rule generation* using decision tree (DT) and Naive Bayes tree (NBTree).

Finance problems solved in this study are churn prediction in bank credit cards customers and fraud detection in insurance. Based on sensitivity measure, the empirical results suggest that the proposed modified active learning-based rule extraction approach yielded best sensitivity and, length and number of rules is reduced resulting in improved comprehensibility. Feature selection leads to the most important attributes of the customers and extracted rules serves as early warning system to the management to enforce better CRM practices and detect/avoid possible frauds.

Keywords: support vector machine; SVM; rule extraction; modified active learning-based approach; insurance fraud; customer relationship management; CRM; normal distribution function; logistic distribution function; customer churn.

Reference to this paper should be made as follows: Farquad, M.A.H., Ravi, V. and Raju, S.B. (2012) 'Analytical CRM in banking and finance using SVM: a modified active learning-based rule extraction approach', *Int. J. Electronic Customer Relationship Management*, Vol. 6, No. 1, pp.48–73.

Biographical notes: M.A.H. Farquad is a Research Assistant at the School of Business, The University of Hong Kong. He holds a PhD in Computer Science from University of Hyderabad, India and Institute for Development and Research in Banking Technology (IDRBT), Hyderabad. His research interests include; data mining, soft computing, banking, finance, and customer relationship management (CRM). He is an Ad Hoc Referee for Information Sciences, Knowledge Based Systems Journal and various IEEE International Conferences. He is a programme committee member of International Conference on Data Mining 2011, Las Vegas and also a technical committee member of 3rd International Conference on Computer Technology and Development, China.

V. Ravi is an Associate Professor in IDRBT, Hyderabad since February 2010. He received his PhD in Soft Computing from Osmania University, Hyderabad and RWTH Aachen, Germany. Earlier, he was a Faculty at NUS, Singapore for three years. He published 105 papers in refereed journals/conferences and invited book chapters. He edited Advances in Banking Technology and Management: Impacts of ICT and CRM published by IGI Global, USA. He is a Referee for several international journals and on the editorial board of *IJIDS*, *IJDATS*, *IJISSS*, *IJWS* and *IJITPM*. He is listed in the *Marquis Who's Who* in the World in 2009/2010 and in Engineering in 2011.

S. Bapi Raju received his BTech in EE from Osmania University, India, MS and PhD from University of Texas at Arlington, USA. He has over ten years of teaching and research experience in neural networks, machine learning and artificial intelligence and their applications. Currently, he is a Professor in the Department of Computer and Information Sciences, University of Hyderabad. He has over 50 publications (journal/conference). His main research interests include biological and artificial neural networks, neural and cognitive modelling, machine learning, pattern recognition, neuroimaging and bioinformatics.

1 Introduction

Customer relationship management (CRM) is a process or methodology used to learn more about customers' needs and behaviours in order to develop stronger relationship with them. CRM involves the continuous use of refined information about current and potential customers in order to anticipate and respond to their needs and draws on a combination of business process and information technology to discover the knowledge about the customers. Questions like, 'who are the new customers?', 'what do they do?' and 'what do they like?' are answered efficiently using effective CRM (Shaw et al., 2001; Chan and Ip, 2011; Wong et al., 2012). Therefore, the effective management of information is critical to the concept of CRM for;

- product tailoring and service innovation (websites tailored to customer needs, taste experience and the development of mass customisation)
- providing a single and consolidated view of the customer
- calculating the lifetime value of the customer
- designing and developing personalised transactions
- communicating with customers using multiple channels
- cross-selling/up-selling various products to customers.

Strategic CRM, *operational* CRM and *analytical* CRM are the three basic types of CRM proposed in literature (Kotler, 2002; Bull, 2010; Adebanjo, 2008; Bhatnagar and Ranjan, 2010; Faase et al. 2011). *Strategic* CRM is focused on developing a customer-centric business culture that is dedicated towards winning and keeping the customers by creating and delivering value better than competitors. *Operational* CRM is focused on the automation of the customer-focusing parts of businesses, such as, marketing, selling and service functions. *Analytical* CRM is concerned with exploiting customer data to enhance both customer and company value and is built upon the foundation of customer information. Analytical CRM can deliver better, timely and even personally customised solutions to the customers' problems, thereby enhancing customer satisfaction. In its simplest form a CRM implementation may be a 'frequently asked questions' (FAQs) page on the company's website, a call-centre, an e-mail newsletter or a customer application download facility (Muther, 2002; Ojasalo, 2002), which facilitate the organisation to do better and customised services to the customers.

In this research work, we proposed modified active learning-based approach (MALBA) to extract rules from the trained SVM model by making use of key concepts of the SVM: the support vectors. Active learning implies the focus on SVM's decision boundary where most of the noise is found, this leads to the poor performance of the rule induction techniques. By generating extra data close to these support vectors that are provided with a class label by the trained SVM model, rule induction techniques are better able to discover suitable rules. The proposed approach in this article is advancement to the earlier study of Farquad et al. (2010b). They used uniform distribution to generate extra instances near support vectors based on the distance between the training instances and support vectors. Later, using Naive Bayes tree (NBTree) rules were generated. In this research study, we employed:

1. feature selection using SVM-RFE
2. normal and logistic distribution functions are applied for synthetic data generation purposes
3. rules are generated using NBTree and decision tree (DT) separately
4. applications analysed are churn prediction in bank credit card customers and insurance fraud detection.

Details of the proposed approach are provided in the proposed approach Section 4 below.

This paper is structured as follows: In Section 2 related works of rule extraction from SVM are presented. Section 3 provides the details about two most important applications of finance that are analysed in the current study. In Section 4, the proposed eclectic rule extraction approach is presented in detail. Section 5 presents the empirical analysis and finally Section 6 concludes this paper.

2 Related work

Although SVM (Vapnik, 1995) generally predicts better, they are *black box* models, i.e., the knowledge learnt by SVM during training is not interpretable by the user. Many researchers tried to treat this *accuracy vs. comprehensibility* trade-off by converting the *black box*, high accurate models to transparent models via *rule extraction* (Gallant, 1988). A learning system might discover salient features in the input data whose importance was not previously recognised (Tickle et al., 1998).

Extensive work has been done towards developing rule extraction techniques for neural networks (Tickle et al., 1998) but less work is reported for extracting rules from SVM (Barakat and Bradley, 2010). Nunez et al. (2002) proposed a rule extraction approach, i.e., SVM + Prototype, where using k-means clustering algorithm and support vectors an ellipsoid is defined in the input space. Later, these ellipsoids are mapped to *if-then* rules. Fung et al. (2005) proposed a rule extraction technique similar to SVM + Prototype but does not include computationally expensive clustering. Instead, the algorithm transforms the problem into a simpler and equivalent variant and constructs hyper cubes by solving linear programming problems. Each hypercube is then transformed into an *if-then* rule.

RulExtSVM (Fu et al., 2004) is proposed for extracting *if-then* rules using intervals defined by hyperrectangular forms, which are generated using the intersection of the support vectors with the decision boundary. Hyperrectangle rules extraction (HRE) (Zhang et al., 2005) first constructs hyperrectangles according to the prototypes and the support vectors, then these hyperrectangles are projected onto coordinate axes and *if-then* rules are formed. Chaves et al. (2005) proposed a decompositional fuzzy rule extraction (FREx) approach, which applies triangular fuzzy membership function to every attribute in the input space and determines the projection of the support vectors in the coordinate axes and fuzzy *if-then* rules are obtained.

Barakat and Bradely (2007) proposed modified sequential covering algorithm termed SQRex-SVM is used to directly extract the rules from support vectors. A multiple kernel-support vector machine (MK-SVM) (Chen et al., 2007) scheme is proposed for feature selection, rule extraction and prediction modelling. It is reported that rules extracted using MK-SVM improves the explanation capacity of SVM and extracts more generalised rules.

Meanwhile the idea of researchers' looking at rule extraction approaches is magnified and they started making use of the SVM as an oracle to modify the data according to its prediction. Later, this modified data is used with rule generating algorithm to generate rules. Such approaches are called as eclectic approaches (Barakat and Diederich, 2004, 2005; Farquad et al., 2009, 2010c, 2010d). Then, Farquad et al. (2008, 2010a) propose a rule extraction approach using only support vectors extracted by developed SVM model.

They employed fuzzy rule-based system to extract fuzzy rules and concluded that more human comprehensible rules are generated.

Recently, active learning-based approach (ALBA) to extract rules from SVM models is proposed by Martens et al. (2009). ALBA makes use of the support vectors to generate additional samples near them. They reported that more data near support vectors and SVM prediction leads to the noise free data which in-turn improves the discrimination power of the rule induction algorithms. Then, Farquad et al. (2010b) proposed modified ALBA for extracting rules from SVM. They suggested the use of support vector set instead of the training set (Martens et al., 2009) to be used along with the generated data. They concluded that, their proposed approach is able to produce less number of rules, increasing the comprehensibility of the extracted rules. In this paper, advancement to Farquad et al. (2010b) is presented.

3 Financial applications analysed

3.1 Churn prediction in bank credit card customers

The problem of customers shifting loyalties from one organisation to another is called ‘churn’, and is common nowadays, which motivated the service industries like banks and insurance to provide better services to their customers. Churn occurs due to various reasons, such as availability of latest technology at the organisation, customer-friendly staff and proximity of geographical location, etc. In the recent past, intelligent techniques have proved to provide insight into the historical data of the organisation and provide the early warnings to the management to be focused on credit worthy customers and they can take precautionary steps to retain such customers. Hence, there is a pressing need to develop algorithmic models that can predict which existing ‘loyal’ customer is going to churn out to the competitor in near future (Kumar and Ravi, 2008).

The churn prediction in bank credit card customers is taken 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. This dataset is obtained from Business Intelligence Cup (2004) and the attribute information is presented in Table 1. This dataset consists of 14814 instances; of which 13812 are loyal customers, i.e., 93% and 1002 are churners, i.e., 7%. Hence, the dataset is highly unbalanced in terms of the proportion of churners versus non-churners.

Over a decade, researchers have applied machine learning techniques for churn prediction problem; such as logistic regression (Bolton et al., 2005; Nie et al., 2011), neural networks (Mozer et al., 2000; Hayashi et al., 2010), random forest (Larivière and Van den Poel, 2004), DT (Farvaresh and Sepehri, 2011), fuzzyARTMAP (Naveen et al., 2009) and SVM (Xia and Jin, 2008; Yu et al., 2011; Huang et al., 2012) are some of the approaches.

3.2 Fraud detection in insurance

The dataset obtained from Pyle (1999) is also highly unbalanced with 94% legitimate and 6% fraudulent customers’ data. This is the only available fraud detection dataset in automobile insurance and it is provided by Angoss Knowledge Seeker software. Originally named ‘carclaims.txt’, it can be found in the accompanying compact disc from

Pyle (1999). This dataset contains 11,338 records from January 1994 to December 1995, and 4,083 records from January 1996 to December 1996. It has a 6% fraudulent and 94% legitimate instances with an average of 430 claims per month. The original dataset has six numerical attributes and 25 categorical attributes, including the binary class label (fraud or legal). Prior to its analysis, pre-processing is carried out to make this dataset feasible for this research study.

Table 1 Attribute information of churn prediction dataset

#	*Feature*	*Description*	*Value*
	Target	Target variable	0 – Non-churner 1 – Churner
1	CRED_T	Credit in month T	Positive real number
2	CRED_T-1	Credit in month T-1	Positive real number
3	CRED_T-2	Credit in month T-2	Positive real number
4	NCC_T	Number of credit cards in months T	Positive integer value
5	NCC_T-1	Number of credit cards in months T-1	Positive integer value
6	NCC_T-2	Number of credit cards in months T-2	Positive integer value
7	INCOME	Customer's income	Positive real number
8	N_EDUC	Customer's educational level	1 – University student 2 – Medium degree 3 – Technical degree 4 – University degree
9	AGE	Customer's age	Positive integer
10	SX	Customers sex	1 – Male 0 – Female
11	E_CIV	Civilian status	1– Single 2 – Married 3 – Widow 4 – Divorced
12	T_WEB_T	Number of web transaction in months T	Positive integer
13	T_WEB_T-1	Number of web transaction in months T-1	Positive integer
14	T_WEB_T-2	Number of web transaction in months T-2	Positive integer
15	MAR_T	Customer's margin for the company in months T	Real number
16	MAR_T-1	Customer's margin for the company in months T-1	Real number
17	MAR_T-2	Customer's margin for the company in months T-2	Real number
18	MAR_T-3	Customer's margin for the company in months T-3	Real number
19	MAR_T-4	Customer's margin for the company in months T-4	Real number
20	MAR_T-5	Customer's margin for the company in months T-5	Real number
21	MAR_T-6	Customer's margin for the company in months T-6	Real number

Table 2 Attribute information of the Insurance data used

#	*Feature*	*Description*
1	Month	Month in which accident took place
2	Week of month	Accident week of month
3	Day of week	Accident day of week
4	Month claimed	Claim month
5	Week of month claimed	Claim week of month
6	Day of week claimed	Claim day of week
7	Year	1994, 1995 and 1996
8	Make	Manufacturer of the car (19 companies)
9	Accident area	Rural or urban
10	Gender	Male or female
11	Marital status	Single, married, widow and divorced
12	Age	Age of policy holder
13	Fault	Policy holder or third party
14	Policy type	Type of the policy (1 to 9)
15	Vehicle category	Sedan, sport or utility
16	Vehicle price	Price of the vehicle with six categories
17	Rep. number	ID of the person who process the claim (16 ID's)
18	Deductible	Amount to be deducted before claim disbursement
19	Driver rating	Driving experience with four categories
20	Days: policy accident	Days left in policy when accident happened
21	Days: policy claim	Days left in policy when claim was filed
22	Past number of claims	Past number of claims
23	Age of vehicle	Vehicle's age with eight categories
24	Age of policy holder	Policy holder's age with nine categories
25	Policy report filed	Yes or no
26	Witness presented	Yes or no
27	Agent type	Internal or external
28	Number of supplements	Number of supplements
29	Address change claim	No. of times change of address requested
30	Number of cars	Number of cars
31	Base policy (BP)	All perils, collision or liability
32	Class	Fraud found (yes or no)

3.2.1 Preprocessing

It is observed that the *age* attribute in the dataset appeared twice in numerical and categorical form as well (Attributes 12 and 24 in Table 2). Hence, the age attribute with numerical values is removed from the data to reduce the complexity caused by too many unique values it possesses. Further, the attributes *year*, *month*, *week of month* and *day of week* represent the date of the accident (Attributes 7, 1, 2 and 3 in Table 2) and the attributes *month claimed*, *week of month claimed* and *day of week claimed* (Attributes 4, 5 and 6 in Table 2) represent the date of the insurance claim. Thus, a new attribute *gap* is derived from seven attributes such as *year*, *month*, *week of month*, *day of week*, *month claimed*, *week of month claimed* and *day of week claimed*. The attribute *gap* represents the time difference between the accident occurrence and insurance claim. Thus, 24 variables which included some derived variables are selected for further study. Hence, we have 15,420 samples with 24 predictor variables and 1 class variable. The dataset under consideration consists of 14,497 instances representing the behaviour of legitimate customers, whereas only 923 instances represent fraudulent customers. Hence, the dataset is highly unbalanced with 94% legitimate instances and 6% fraudulent instances. The attribute information of the actual dataset and pre-processed dataset are presented in Table 2 and Table 3, respectively.

Table 3 Attribute information of the pre-processed insurance fraud data

#	*Feature*	*Description*
1	Gap	Time difference of accident and insurance claim
2	Make	Manufacturer of the car (19 companies)
3	Accident area	Rural or urban
4	Gender	Male or female
5	Marital status	Single, married, widow and divorced
6	Fault	Policy holder or third party
7	Policy type	Type of the policy (1 to 9)
8	Vehicle category	Sedan, sport or utility
9	Vehicle price	Price of the vehicle with six categories
10	Rep. number	ID of the person who process the claim (16 ID's)
11	Deductible	Amount to be deducted before claim disbursement
12	Driver rating	Driving experience with four categories
13	Days: policy accident	Days left in policy when accident happened
14	Days: policy claim	Days left in policy when claim was filed
15	Past number of claims	Past number of claims
16	Age of vehicle	Vehicle's age with eight categories
17	Age of policy holder	Policy holder's age with nine categories
18	Policy report filed	Yes or no
19	Witness presented	Yes or no
20	Agent type	Internal or external
21	Number of supplements	Number of supplements
22	Address change claim	No of times change of address requested
23	Number of cars	Number of cars
24	Base policy (BP)	All perils, collision or liability
25	Class	Fraud found (yes or no)

Intelligent techniques proposed for solving fraud detection problem are principal component analysis (Brockett et al., 2002), AdaBoosted Naïve Bayes (Viaene et al., 2004), stacking using neural network, Naive Bayesian and DT (Phua et al., 2004), adaptive fraud detection (Fawcett and Provost, 1997) and so on. Wheeler and Aitken (2000), and Phua et al. (2010) have explored multiple classification techniques for fraud detection.

4 Proposed rule extraction approach

In this research work, we proposed modified ALBA to extract rules from the trained SVM model by making use of key concepts of the SVM: the support vectors. Active learning implies the focus on SVM's decision boundary where most of the noise is found, this leads to the poor performance of the rule induction techniques. By generating extra data close to these support vectors that are provided with a class label by the trained SVM model, rule induction techniques are better able to discover suitable rules. The proposed approach in this article is advancement to the earlier study of Farquad et al. (2010b). They used uniform distribution to generate extra instances near support vectors based on the distance between the training instances and support vectors. Later, using NBTree rules were generated. In this research study, we employed:

1 feature selection using SVM-RFE

2 normal and logistic distribution functions are applied for synthetic data generation purposes

3 rules are generated using NBTree and DT separately

4 applications analysed are churn prediction in bank credit card customers and insurance fraud detection.

The proposed approach comprises three phases; *feature selection phase*, *active learning phase* and *rule generation phase*. During *feature selection phase*, feature selection using SVM-RFE is carried out. Later, during active learning phase, full feature data and reduced feature data is used and synthetic data is generated near support vectors for which predictions are obtained using developed SVM model. Finally, during *rule generation phase*, NBTree and DT algorithms are employed separately on modified datasets to generate rules. The architecture of the proposed approach is depicted in Figure 1.

4.1 Active learning phase of the proposed approach

Let training data be denoted by D_{tr} and the number of instances in D_{tr} by N_{tr}.

Step 1: Train SVM and obtain the support vectors using D_{tr}.

/* Calculate the average distance $distance_k$ of training data to support vectors, in each dimension k */

Step 2:

for k = 1 to n do

    $distance_k = 0$

for all support vectors SV_j do

for all training data instance d in D_{tr} do

$distance_k = distance_k + | d_k - SV_{j,k} |$

end for

end for

$distance_k = \dfrac{distance_k}{\# SVs + N_{tr}}$, where #SVs is the number of support vectors.

end for

/* *Modified ALBA (Proposed)* */

Step 3 Randomly generate an extra data instance x_i following uniform distribution [–1, 1] and close to support vectors

For i = 1 to 500 or 1000 do

For k = 1 to n do

$x_{i,k} = sv(j,k) + \left[(2 * rand - 1) \times \dfrac{disance_k}{2} \right]$ with *rand* a random number in [0, 1]

End for

/* Append the generated data to the support vectors */

Step 4 Provide a class label y_i using the trained SVM as oracle.

End for

For generating extra instances near support vectors, we employed Normal and Logistic distribution function separately i.e. step 3 of active learning phase.

Normal distribution function (Box Muller approach)

for k = 1 to n do

for i = 1 to 500 or 1,000 do

$y_{i,k}$ = sqrt(–2 × log($x_{i,k}$)) × cos(2 × Π × ($x_{i+1,k}$));

$y_{i+1,k}$ = sqrt(–2 × log($x_{i,k}$)) × sin(2 × Π × ($x_{i+1,k}$));

end for

end for

Logistic distribution function

for i = 1 to 500 or 1,000 do

for k = 1 to n do

$x_{i,k}$ = rand;

end for

end for

for i =1 to 500 or 1,000

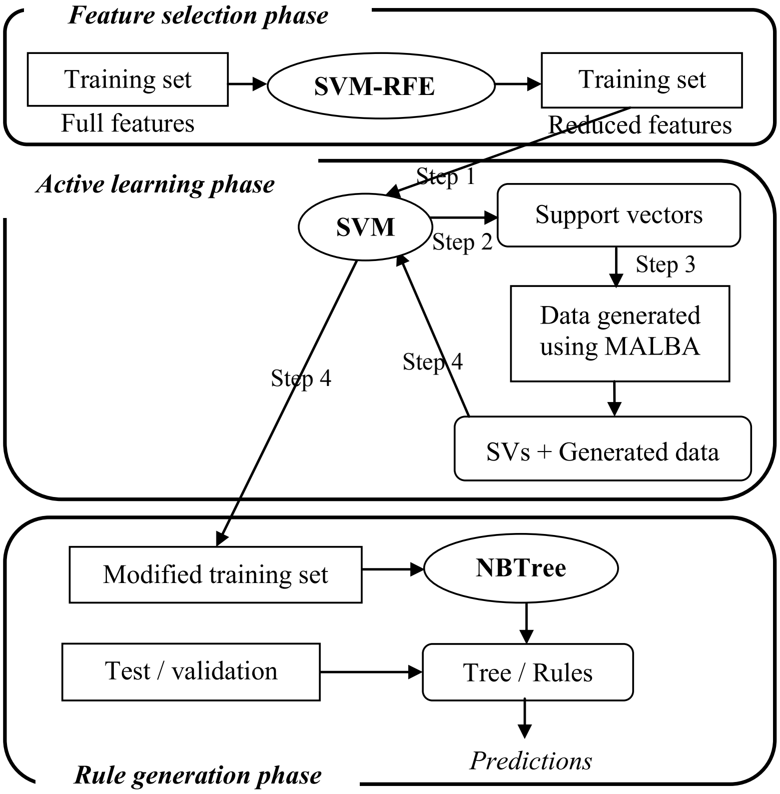
for k = 1 to n do

$u_{i,k}$ = 1 / (1 + exp(–$x_{i,k}$));

end for

end for

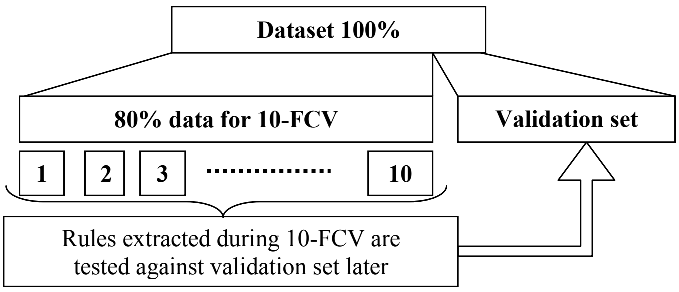
Figure 1 Architecture of the proposed rule extraction approach



4.2 Experimental setup

The available medium scale unbalanced dataset is first divided into two parts of 80:20 ratios. 80% of the data is then used for training under ten fold cross-validation (10-FCV) method using stratified random sampling and 20% data is stored untouched for validation purpose later. This 20% data represents the reality and originality present in the original data used for evaluating the efficiency of the rules generated under 10-FCV. The experimental setup followed in this paper is depicted in Figure 2.

Figure 2 Experimental setup followed



To evaluate and compare the efficiency of the rules extracted using proposed approach, five different cases are investigated as follows;

1 original ALBA (Martens et al., 2009) (range [–0.5 to 0.5], training set)

2 ALBA (SVs) (range [–0.5 to 0.5], support vectors set)

3 MALBA (Farquad et al., 2010d) (range [–1 to 1], support vectors set)

4 MALBA (normal) (normal distribution, support vectors set)

5 MALBA (logistic) (logistic distribution, support vectors set).

Using available training data, SVM model is first developed under 10-FCV and support vectors are extracted for each fold. The distance between the support vectors and training instances is then calculated before generating the extra instances. Based on the previous studies (Martens et al., 2009; Farquad et al., 2010b), 500 and 1,000 extra data instances are generated for empirical analysis in the present study. The generated data is appended to the support vectors set and the predictions are obtained using the trained SVM model and the actual target values are replaced by the predictions of SVM. This modified data is then fed to NBTree (Kohavi, 1996) and DT (Quinlan, 1986) to generate rules. The process flow of the proposed approach is presented in Figure 1, feature selection is invoked when it is employed else it is not invoked when full feature data is used.

5 Results and discussions

Identifying potential churners and possible frauds correctly is the basic intension of many business decision makers. Hence, they place high emphasis on sensitivity measure alone which contributes towards the bottom-line of the fundamental CRM. Consequently, in this paper, sensitivity is accorded top priority ahead of specificity and accuracy. We used Rapidminer tool (Mierswa et al., 2006) for SVM model building, support vector extraction, NBTree and DT (J48). Active learning step is implemented in MATLAB. The quantities employed to measure the quality of the classifiers are sensitivity, specificity and accuracy, which are defined as follows (Fawcett, 2006).

Sensitivity is the measure of proportion of the true positives (churn or fraud in this study), which are correctly identified.

$$Sensitivity = \frac{\text{True positive}}{(\text{True positive} + \text{False negative})}$$

Specificity is the measure of proportion of the true negatives (loyal or legitimate in this study), which are correctly identified.

$$Specificity = \frac{\text{True negative}}{(\text{True negative} + \text{False positive})}$$

Accuracy is the measure of proportion of true positives and true negatives, which are correctly identified.

$$Accuracy = \frac{\text{True positive} + \text{True negative}}{(\text{True Positive} + \text{True negative} + \text{False positive} + \text{False negative})}$$

A rule set is considered to display a high level of *fidelity* if it can *mimic* the behaviour of the machine learning technique from which it was extracted, i.e., SVM in our study. Based on sensitivity, the classifiers are compared using t-test at n1 + n2 – 2 = 10 + 10 – 2 = 18 degrees of freedom at 10% level of significance. We tested if the difference in performances is statistically significant. The tabulated value of t-test for 18 degrees of freedom at 10% level of significance is 1.73. That means, if the computed t-test value between two different classifiers is more than 1.73, then we can say that the difference between techniques is statistically significant and otherwise not significant.

The results presented in Tables 4 through 27 indicate the average sensitivity, specificity, accuracy and t-test values under 10-FCV and against validation set. Last two columns in the results, table show the time taken and the number of rules extracted using various cases discussed in the previous Section 4.2. Tables 4 through 9 show the average results obtained for churn prediction in bank credit card data using rules extracted of NBTree including average fidelity table and best rule set table. It is observed that feature selection in first phase certainly improves the efficiency of the rules with respect to sensitivity, fidelity and comprehensibility of the rules. It is concluded here that feature selection using SVM-RFE and synthetic data generation near support vectors leads to improved efficiency of the rule induction algorithm.

Table 4 Average results of churn prediction using SVM + NBTree (500 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
SVM	63.35	81.41	80.19	4.928	64.65	80.63	79.55	4.786	7.9	
ALBA	65.48	85.28	83.92	3.436	67.7	84.52	83.38	3.125	39.1	31.8
ALBA (SVs)	74.93	83.05	82.5	0.82	76.55	82.74	82.32	0.755	14.7	13.3
MALBA	*78.17*	*80.36*	*80.28*	-	*79.35*	*79.16*	*79.17*	-	*5.2*	*11.1*
MALBA (normal)	77.8	81.91	81.63	0.108	78.45	81.63	81.4	0.288	14.5	14.3
MALBA (logistic)	74.44	80.38	79.98	0.948	74.05	80.18	79.75	1.445	12.8	19

Table 5 Average results of churn prediction using SVM + NBTree (1,000 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
ALBA	65.36	85.88	84.51	2.855	68.05	84.75	83.62	2.407	52	27.6
ALBA (SVs)	73.8	83.26	82.62	0.42	73.25	82.93	82.27	0.978	17.5	13.9
MALBA	74.3	82.84	82.31	0.31	75.9	83.3	82.87	0.239	*6.9*	*12.9*
MALBA (normal)	*75.5*	*82.77*	*82.27*	-	*76.63*	*82.72*	*82.31*	-	*13.7*	*12.7*
MALBA (logistic)	71.45	79.8	79.23	1.01	75.2	79.62	79.53	0.519	12.3	17.6

Table 6 Average results of churn prediction using feature selection + SVM + NBTree (500 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
SVM	85.77	74.45	75.22	0.366	82.85	74.25	74.79	0.849	3.8	
ALBA	*86.53*	*74.81*	*76.10*	-	*84.00*	*74.79*	*75.41*	-	14	46.8
ALBA (SVs)	85.66	75.31	76.01	0.516	82.55	75.13	75.63	1.574	6.3	19.6
MALBA	84.91	75.59	76.22	0.791	83.18	75.77	76.43	0.527	6.3	22.8
MALBA (normal)	85.41	74.76	75.48	0.569	82.30	75.08	75.57	1.156	*6.1*	*20*
MALBA (logistic)	85.29	74.60	75.33	0.632	83.35	74.56	75.15	0.443	*5.8*	*14.9*

Table 7 Average results of churn prediction using feature selection + SVM + NBTree (1,000 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
ALBA	86.53	76.31	75.18	0.105	84.40	74.16	74.85	0.121	16.8	48.1
ALBA (SVs)	86.28	74.35	75.16	0.224	83.30	74.34	74.95	0.972	6.6	24.6
MALBA	*86.77*	*73.25*	*74.16*	-	*84.60*	*73.32*	*74.14*	-	4.9	21.2
MALBA (normal)	84.67	75.15	75.80	0.972	82.15	75.36	75.82	1.799	6.3	21.4
MALBA (logistic)	84.91	75.15	75.79	0.825	82.95	75.36	75.87	1.267	*6*	*16*

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

Table 8 Average fidelity for churn prediction using SVM + NBTree

	Full features		*Feature selection*	
	500	*1,000*	*500*	*1,000*
ALBA	83.28	81.64	95.89	95.89
ALBA (SVs)	80.88	79.03	93.9	93.47
MALBA	82.65	79.1	93.58	91.92
MALBA (normal)	82.16	82.42	93.63	93.59
MALBA (logistic)	80.91	80.36	93.65	94.2

Tables 10 through 15 show the average results obtained using rules extracted of NBTree for insurance fraud detection data, results tables include average fidelity table and best rule set table as well. Similar to earlier findings, it is observed with fraud detection data that feature selection in first phase improves the efficiency of the rules with respect to sensitivity, fidelity and comprehensibility of the rules. It is observed again, that feature selection using SVM-RFE and synthetic data generation near support vectors leads to the improved efficiency of the rule induction algorithm. Empirical results clearly indicate that data generation using normal or logistic distribution function does not improve the results. Selection of any particular distribution function is purely based on data and trial and error basis.

Table 9 Rules extracted for churn prediction using NBTree

#	*Antecedents*	*Consequent*
1	If *CRED-T* <= 594.94 and *CRED-T-2* <= 96.005 and *CRED-T-1* <= 100.12 and *NCC-T* <= 0.5 and *NCC-T-2* <= 0.5 and *T-WEB-T* <= 2.5	Churn
2	If *CRED-T* <= 594.94 and *CRED-T-2* <= 96.005 and *CRED-T-1* <= 100.12 and *NCC-T* <= 0.5 and *NCC-T-2* <= 0.5 and *T-WEB-T* > 2.5	Churn
3	If *CRED-T* <= 594.94 and *CRED-T-2* <= 96.005 and *CRED-T-1* <= 100.12 and *NCC-T* <= 0.5 and *NCC-T-2* > 0.5	Churn
4	If *CRED-T* <= 594.94 and *CRED-T-2* <= 96.005 and *CRED-T-1* <= 100.12 and *NCC-T* > 0.5	Churn
5	If *CRED-T* <= 594.94 and *CRED-T-2* <= 96.005 and *CRED-T-1* > 100.12 and *CRED-T-1* <= 147.9	Churn
6	If *CRED-T* <= 594.94 and *CRED-T-2* <= 96.005 and *CRED-T-1* > 100.12 and *CRED-T-1* > 147.9	Non-churn
7	If *CRED-T* >= 579 and *CRED-T* <= 594.94 and *CRED-T-2* > 96.005 and *T-WEB-T* <= 8.5 and *CRED-T-1* <= 122.5	Churn
8	If *CRED-T* >= 579 and *CRED-T* <= 594.94 and *CRED-T-2* > 96.005 and *T-WEB-T* <= 8.5 and *CRED-T-1* > 122.5	Non-churn
9	If *CRED-T* >= 579 and *CRED-T* <= 594.94 and *CRED-T-2* > 96.005 and *T-WEB-T* <= 8.5	Churn
10	If *CRED-T* <= 594.94 and *CRED-T-2* >= 96.005 and *CRED-T-2* <= 105.2 and *T-WEB-T* > 8.5	Non-churn
11	If *CRED-T* > 594.94	Non-churn

Table 10 Average results of insurance fraud detection using SVM + NBTree (500 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
SVM	71.13	63.98	64.40	1.2	70.76	63.20	63.65	3.554	6.8	
ALBA	71.42	63.61	64.07	1.071	70.65	62.68	63.16	3.082	51.3	29
ALBA (SVs)	73.05	60.91	61.61	0.453	74.70	60.46	61.34	0.537	13.5	10.9
MALBA	*74.01*	*60.40*	*61.21*	-	*75.73*	*59.76*	*60.62*	-	*6.2*	*10*
MALBA (normal)	72.23	60.89	61.57	0.875	73.41	59.91	60.72	1.793	16.1	18
MALBA (logistic)	72.65	60.41	60.20	0.519	74.35	59.24	60.15	0.873	21.1	16

Table 11 Average results of insurance fraud detection using SVM + NBTree (1,000 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
ALBA	70.08	64.18	64.53	1.232	70.22	62.74	63.19	1.575	62.3	44
ALBA (SVs)	*73.06*	*59.67*	*60.46*	-	74.11	58.78	59.42	0.236	18	16.1
MALBA	71.41	55.01	55.98	0.452	*75.21*	*54.69*	*55.97*	-	*13.2*	*15*
MALBA (normal)	72.53	59.91	60.67	0.172	74.29	59.15	60.06	0.271	27	28
MALBA (logistic)	72.23	59.92	60.71	0.346	75.05	58.33	59.84	0.04	23.9	17

Table 12 Average results of insurance fraud detection using feature selection + SVM + NBTree (500 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
SVM	84.24	57.07	58.76	0.25	87.68	56.36	58.21	0.209	3.7	
ALBA	84.37	56.91	58.56	0.215	88.00	56.27	58.17	0.118	12.6	30
ALBA (SVs)	*85.32*	*55.73*	*57.50*	-	*88.43*	*55.13*	*57.19*	-	*5.8*	*12.2*
MALBA	84.41	56.28	57.97	0.214	88.22	55.64	57.59	0.059	*2.5*	*13*
MALBA (normal)	84.64	56.36	58.06	0.156	88.00	55.74	57.67	0.119	8.6	16
MALBA (logistic)	84.65	55.20	56.97	0.161	88.16	54.77	56.73	0.074	9.8	15

Table 13 average results of insurance fraud detection using feature selection + SVM + NBTree (1,000 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
ALBA	83.70	57.18	58.76	0.343	87.84	56.56	58.44	0.177	12.9	26
ALBA (SVs)	84.64	55.96	57.68	0.123	88.05	55.33	57.29	0.115	6.3	15.3
MALBA	*85.18*	*56.22*	*57.96*	-	*88.48*	*55.60*	*57.57*	-	*3.5*	*17*
MALBA (normal)	84.24	56.23	57.91	0.215	88.11	55.88	57.82	0.103	9.8	18
MALBA (logistic)	85.05	56.08	57.82	0.027	88.38	55.34	57.34	0.028	10.6	16

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

Table 14 Average fidelity for insurance fraud detection using SVM + NBTree

	Full features		*Feature selection*	
	500	*1,000*	*500*	*1,000*
ALBA	92.78	92.81	99.47	99.5
ALBA (SVs)	89.22	88.25	96.93	97.09
MALBA	88.82	88.97	97.08	96.69
MALBA (normal)	88.13	87.78	97.27	97.48
MALBA (logistic)	87.87	88.4	97.04	96.77

Table 15 Rules extracted for insurance fraud detection using NBTree

#	*Antecedents*	*Consequent*
1	If *marital status* is single and *base policy* is all perils and *age of vehicle* is less than six years then	Non-fraud
2	If *marital status* is single and *base policy* is collision/liability then	Fraud
3	If *marital status* is single and *base policy* is all perils and *age of vehicle* is more than six years then	Non-fraud
4	If *marital status* is married/widow/divorced and *manufacturer* is top 9 from manufacturers list and *vehicle category* is Sedan and *fault* is of policy holder	Fraud
5	If *marital status* is married/widow/divorced and *manufacturer* is top 9 from manufacturers list and *vehicle category* is Sedan and *fault* is of third party	Non-fraud
6	If *marital status* is married/widow/divorced and *manufacturer* is top 9 from manufacturers list and *vehicle category* is sports/utility and *age of vehicle* is more than three years	Fraud
7	If *marital status* is married/widow/divorced and *manufacturer* is after 9th in the list	Fraud

Table 16 Average results of churn prediction using SVM + DT (500 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
SVM	63.35	81.41	80.19	3.166	64.65	80.63	79.55	5.414	7.9	
ALBA	60.22	83.56	81.97	3.943	65.2	82.65	81.48	4.541	37.6	117.3
ALBA (SVs)	69.68	80.38	79.66	1.296	72.7	80.38	79.25	1.725	10.7	92.9
MALBA	73.81	76.91	76.7	0.361	75.05	75.99	76.43	0.914	10.4	85.5
MALBA (normal)	*75.42*	*82.97*	*82.45*	-	*77.25*	*82.64*	*82.28*	-	10.3	46.8
MALBA (logistic)	71.04	83.28	82.46	0.767	75.2	83.19	82.71	0.531	*9.9*	*29*

DT is one of the most popular rule induction algorithms. Average results obtained for churn prediction data using rules extracted of DT are presented in Table 16 through Table 21 including average fidelity table and best rule set. Similar to the results of NBTree rules, it is observed that feature selection using SVM-RFE and synthetic data generation near support vectors leads to the improved efficiency of the rule induction algorithm, i.e., DT. It is also observed that using normal distribution function to generated synthetic data is a better option when full feature data is considered.

Table 17 Average results of churn prediction using SVM + DT (1,000 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
ALBA	61.6	83.81	82.32	2.875	64.05	83	81.72	4.123	40.3	127
ALBA (SVs)	67.7	80.55	79.68	1.031	71.65	79.71	79.17	1.679	10.4	109.5
MALBA	70.1	81.6	80.85	0.481	72.85	81.3	80.7	1.432	11.1	96.5
MALBA (normal)	*72.04*	*83.58*	*82.35*	-	76.1	82.91	82.45	0.337	11	54.9
MALBA (logistic)	71.81	80.74	80.14	0.054	*77.2*	*80.58*	*80.35*	-	*10.2*	*32.9*

Table 18 Average results of churn prediction using feature selection + SVM + DT (500 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
SVM	85.77	74.45	75.22	0.11	82.85	74.25	74.79	0.206	3.8	
ALBA	84.91	74.77	75.45	0.308	81.90	74.59	75.09	1.167	7.3	77.8
ALBA (SVs)	*85.54*	*75.20*	*75.90*	-	*83.05*	*75.46*	*75.97*	-	5.8	25
MALBA	85.03	76.19	76.79	0.226	82.35	76.44	76.84	0.655	5.8	30.1
MALBA (normal)	85.29	75.86	76.47	0.117	82.30	76.36	76.76	0.794	5.8	26.2
MALBA (logistic)	85.53	75.37	76.06	0.005	82.85	75.50	76.00	0.211	*5.8*	*18.4*

Table 19 Average results of churn prediction using feature selection + SVM + DT (1,000 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
ALBA	85.65	74.65	75.40	0.576	81.75	74.50	74.99	2.031	7.6	93.1
ALBA (SVs)	*86.77*	*71.25*	*72.36*	-	*83.40*	*71.47*	*72.26*	-	5.8	36.1
MALBA	84.90	75.05	75.72	0.214	82.89	75.32	75.83	0.061	5.78	40.78
MALBA (normal)	85.65	75.01	75.73	0.568	82.00	75.41	75.85	1.46	5.8	39.1
MALBA (logistic)	85.16	75.23	75.90	0.793	82.55	75.41	75.89	1.078	*5.8*	*19.5*

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

Table 20 Average fidelity for churn prediction using SVM + DT

	Full features		*Feature selection*	
	500	*1,000*	*500*	*1,000*
ALBA	84.46	85.66	96.78	96.9
ALBA (SVs)	82.49	83.01	92.95	91.39
MALBA	83.46	83.7	94.03	93.19
MALBA (normal)	83.73	84.37	93.94	93.87
MALBA (logistic)	82.24	83.15	93.967	94.11

Table 21 Rules generated for churn prediction using DT

#	*Antecedents*	*Consequent*
1	If *CRED-T* <= 594.78 and *CRED-T-2* <= 97.81 and *T-WEB-T* <=5	Churner
2	If *CRED-T* <= 594.78 and *CRED-T-2* <= 97.81 and *T-WEB-T* > 5	Non-churner
3	If *CRED-T* <= 594.78 and *CRED-T-2* > 97.81 and *CRED-T-1* <= 92.81	Churner
4	If *CRED-T* <= 594.78 and *CRED-T-2* > 97.81 and *CRED-T-1* > 92.81	Churner
5	If *CRED-T* >= 580.6 and *CRED-T* <= 594.78 and *CRED-T-2* >= 97.81 and *CRED-T-2* <= 102.7 and *CRED-T-1* >= 92.81 and *CRED-T-1* <= 95.31 and *T-WEB-T* <=1	Churner
6	If *CRED-T* >= 580.6 and *CRED-T* <= 594.78 and *CRED-T-2* >= 97.81 and *CRED-T-2* <= 102.7 and *CRED-T-1* > 92.81	Non-churner
7	If *CRED-T* >= 580.6 and *CRED-T* <= 594.78 and *CRED-T-2* >= 97.81 and *CRED-T-2* <= 102.7 and *CRED-T-1* > 92.81 and *T-WEB-T* > 1	Non-churner
8	If *CRED-T* >= 580.6 and *CRED-T* <= 594.78 and *CRED-T-2* >= 97.81 and *CRED-T-2* <= 102.7 and *CRED-T-1* >= 92.81 and *CRED-T-1* <= 95.31 and *T-WEB-T* <=1	Churner
9	If *CRED-T* >= 580.6 and *CRED-T* <= 594.78 and *CRED-T-2* >= 97.81 and *CRED-T-2* <= 102.7 and *CRED-T-1* >= 92.81 and *CRED-T-1* <= 95.31 and *T-WEB-T* <=1	Non-churner
10	If *CRED-T* >= 594.78 and *CRED-T* <= 606.94 and *CRED-T-2* <= 96.96 and *CRED-T-1* <= 91.21	Churner
11	If *CRED-T* >= 594.78 and *CRED-T* <= 606.94 and *CRED-T-2* >= 84.18 and *CRED-T-2* <= 96.96 and *CRED-T-1* > 91.21	Churner
12	If *CRED-T* >= 594.78 and *CRED-T* <= 606.94 and *CRED-T-2* >= 84.18 and *CRED-T-2* <= 96.96 and *CRED-T-1* >= 91.21 *CRED-T-1* <= 97.69 and *T-WEB-T* <= 5	Churner
13	If *CRED-T* >= 594.78 and *CRED-T* <= 606.94 and *CRED-T-2* >= 84.18 and *CRED-T-2* <= 96.96 and *CRED-T-1* >= 91.21 *CRED-T-1* <= 97.69 and *T-WEB-T* > 5	Non-churner
14	If *CRED-T* >= 594.78 and *CRED-T* <= 606.94 and *CRED-T-2* >= 84.18 and *CRED-T-2* <= 96.96 and *CRED-T-1* >= 91.21 *CRED-T-1* <= 97.69	Non-churner
15	If *CRED-T* >= 594.78 and *CRED-T* <= 606.94 and *CRED-T-2* >= 84.18 and *CRED-T-2* <= 96.96 and *CRED-T-1* > 91.21	Non-churner
16	If *CRED-T* >= 594.78 and *CRED-T* <= 606.94 and *CRED-T-2* >= 78.23 and *CRED-T-2* <= 96.96 and *CRED-T-1* <= 95.84	Churner

Table 21 Rules generated for churn prediction using DT (continued)

#	*Antecedents*	*Consequent*
17	If *CRED-T* >= 594.78 and *CRED-T* <= 606.94 and *CRED-T-2* >= 78.23 and *CRED-T-2* <= 96.96 and *CRED-T-1* > 95.84	Non-churner
18	If *CRED-T* >= 594.78 and *CRED-T* <= 606.94 and *CRED-T-2* > 78.23	Non-churner
19	If *CRED-T* > 594.78 and *CRED-T-2* > 96.96	Non-churner

Average results obtained for fraud detection data using rules extracted of DT are presented in Tables 22 through 27 including average fidelity table and best rule set. It is observed from empirical results that feature selection using SVM-RFE and synthetic data generation near support vectors leads to the improved efficiency of the rule induction algorithm. It is also observed that data generation using logistic distribution function with reduced feature set leads to the improvement of DT rules.

Table 22 Average results of insurance fraud detection using SVM+DT (500 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
SVM	71.13	63.98	64.40	1.07	70.76	63.20	63.65	2.033	6.8	
ALBA	71.17	63.80	64.24	0.983	71.24	63.19	63.67	1.565	40.8	187
ALBA (SVs)	73.86	62.01	62.72	0.107	72.76	61.57	62.25	0.752	8.8	50.2
MALBA	*74.14*	*64.61*	*62.36*	-	*74.27*	*61.00*	*62.08*	-	*8.8*	*48*
MALBA (normal)	69.97	63.50	63.89	1.423	71.35	62.87	63.38	1.281	8.8	49
MALBA (logistic)	72.53	61.90	62.53	0.486	72.49	61.39	62.06	0.801	8.8	48

Table 23 Average results of insurance fraud detection using SVM+DT (1,000 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
ALBA	70.21	63.91	64.28	0.567	69.62	63.65	64.00	1.578	45	196
ALBA (SVs)	71.02	63.27	63.73	0.325	70.97	62.25	62.77	0.974	8.8	63.6
MALBA	*72.11*	*60.96*	*60.69*	-	*73.68*	*59.13*	*60.00*	-	*8.8*	*67*
MALBA (normal)	70.76	62.98	63.45	0.383	70.70	62.29	62.80	1.192	8.8	70
MALBA (logistic)	71.03	62.38	62.96	0.302	71.08	61.80	62.36	0.99	*8.8*	*64*

Table 24 Average results of insurance fraud detection using feature selection + SVM + DT (500 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
SVM	84.24	57.07	58.76	0.064	87.68	56.36	58.21	0.086	3.7	
ALBA	84.11	57.20	58.81	0.095	87.78	56.48	58.36	0.056	5.7	37
ALBA (SVs)	84.51	56.49	58.19	0.001	87.84	55.96	57.87	0.042	4.7	16.3
MALBA	83.83	56.71	58.33	0.158	87.67	56.03	57.92	0.089	*4.7*	*17*
MALBA (normal)	84.37	57.15	58.78	0.033	87.62	56.58	58.44	0.101	4.7	21
MALBA (logistic)	*84.51*	*56.74*	*58.31*	-	*87.99*	*56.58*	*58.46*	-	*4.7*	*17*

Table 25 Average results of insurance fraud detection using feature selection + SVM + DT (1,000 extra instances)

	Test				*Validation*				*Time*	*Rules*
	Sens	*Spec*	*Acc*	*t-test*	*Sens*	*Spec*	*Acc*	*t-test*		
ALBA	83.97	57.17	58.77	0.152	87.57	56.48	58.34	0.09	5.7	41
ALBA (SVs)	83.83	56.93	58.56	0.179	87.62	56.37	58.25	0.074	4.7	23.3
MALBA	84.51	56.51	58.16	0.03	*87.89*	*55.81*	*57.74*	-	*4.7*	*24*
MALBA (normal)	83.97	57.87	59.39	0.152	87.77	57.22	59.02	0.032	4.7	26
MALBA (logistic)	*84.64*	*56.90*	*58.55*	-	87.84	56.31	58.20	0.015	*4.7*	*22*

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

Table 26 Average fidelity for insurance fraud detection using SVM + DT

	Full features		*Feature selection*	
	500	*1,000*	*500*	*1,000*
ALBA	91.66	91.68	99.57	99.56
ALBA (SVs)	88.37	88.92	97.71	97.72
MALBA	88.15	85.39	97.81	97.81
MALBA (normal)	88.43	85.88	98.19	98.08
MALBA (logistic)	88.63	88.08	97.89	97.93

Table 27 Rules extracted for insurance fraud detection using DT

#	*Antecedents*	*Consequent*
1	If policy holder is at *fault* and *base policy* is all perils and *vehicle category* is Sedan	Non-fraud
2	If policy holder is at *fault* and *base policy* is all perils and *vehicle category* is sports/utility and age of vehicle is less than 7	Fraud
3	If policy holder is at *fault* and *base policy* is all perils and *vehicle category* is sports/utility and *age of vehicle* is more than 7 and *accident area* is rural and *manufacturer* is top 4 in the list	Non-fraud
4	If policy holder is at *fault* and *base policy* is all perils and *vehicle category* is sports/utility and *age of vehicle* is more than 7 and *accident area* is rural and *manufacturer* is after top 4 in the list	Fraud
5	If policy holder is at *fault* and *base policy* is all perils and *vehicle category* is sports/utility and *age of vehicle* is more than 7 and *accident area* is urban	Fraud
6	If third party is at *Fault*	Non-fraud
7	If policy holder is at *fault* and *base policy* is collision/liability and *vehicle category* is Sedan and *base policy* is all perils/collision and *marital status* is single	Non-fraud
8	If policy holder is at *fault* and *vehicle category* is Sedan and *marital status* is married/widowed/divorced and *manufacturer* is top 3 in the list and *age of vehicle* is less than 6	Fraud
9	If policy holder is at *fault* and *vehicle category* is Sedan and *marital status* is married/widowed/divorced and *manufacturer* is top 3 in the list and *age of vehicle* is more than 6	Non-fraud
10	If policy holder is at *fault* and *vehicle category* is Sedan and *marital status* is married/widowed/divorced and *manufacturer* is after top 3 in the list	Fraud
11	If policy holder is at *fault* and *base policy* is collision/liability and *vehicle category* is spots/utility	Fraud
12	If policy holder is at *fault* and *base policy* is collision/liability and *vehicle category* is Sedan and *base policy* is liability	Fraud

5 Overall observations at a glance

It is observed from empirical results that the hybrids with extra instances combined with support vector set perform better than the original ALBA of Martenes et al. (2009), where generated instances were combined with training set. Generated instances with training set results in increased number of instances and add to the complexity of the rule induction algorithm. It is also observed that the time taken and the number of rules extracted using proposed rule extraction approach is very much less compared to the original ALBA of Martenes et al. (2009). It is observed that the proposed approach performed better with reduced feature data instead of full feature data with respect to sensitivity, number of rules, size of the rules and the time taken for rule extraction. It is also observed that rules generated using NBTree perform better compared to that of the rules extracted using DT with respect to accuracy, time taken, number of rules and rule length. Possible reason for such behaviour of NBTree could be the tree generated by

NBTree, where leaf node, instead of having prediction for a class it has probability for the number of classes in training set. The only drawback of the proposed approach is the selection of distribution function to be employed for generating extra data, as it is observed that no distribution function is consistent towards improvement of results.

Through management's perspective, the generated rule set show that, for churn prediction problem it is credit in the account during current month and previous two months are the main driving element to determine customer churn. Based on the generated rules, it is also observed that, web transactions during current month and previous month also sometimes helps in deciding about to churn customers, likewise management can take precautionary steps to avoid such churns. Credit card transactions in current month also plays an important role to understand about to churn customers. If any of these attributes count is less than the average user count then management has to think over and has to come out with better policies to improve these counts and retain the existing customer. Instead of looking at all the attributes of the customers', the feature set presented here are well enough for the management to understand about the customers' behaviour.

For fraud detection problems based on the extracted rules, it is observed that marital status and base policy are the main driving elements for frauds. It is also observed from the rules that most the time unmarried/single customers are tend to be fraud. Rules also show that most of the accidents, claims and frauds are about sports or utility vehicles only, whereas claims for Sedan category vehicles are tend to be non-frauds. It is also observed that when the accident took place because of the third party instead of the policy holder it is tend to be a non-fraud case, whereas if the fault is of the policy holder then there may be chance of fraud. Heavy losses can be avoided by insurance industry if they can predict the possible fraud prior to its occurrence. Considering these attributes insurance company can make prior predictions about the authenticity of claim and can accept the claim or investigate further about the claim.

6 Conclusions

In this paper, we present a novel and extended modified ALBA for rule extraction from SVM to mine two unbalance medium scale data mining problems such as; churn prediction in bank credit cards customers and Insurance fraud detection. Feature selection in employed in first phase of the proposed approach and active learning is carried out during second phase, where extra data is generated near support vectors and the predictions are obtained using the trained SVM model. Finally, rule induction algorithms viz., NBTree and DT are employed separately to generate rules. Based on the sensitivity measure, it is observed that the proposed rule extraction approach using reduced features yielded better sensitivity compared to that of the sensitivity with full features. Because of the feature selection in the first phase, the number of rules and the size of the rules extracted are small, resulting in the improvement of comprehensibility of the system. This justifies our claim to generate extra instances near support vectors to improve the discrimination power of rule induction algorithms. The only drawback observed in this research study is the selection of the distribution function to generate the extra data near support vectors and it is purely depend on trial and error basis.

References

Adebanjo, D. (2008) 'The application of process modelling and simulation techniques to analysing the impact of CRM implementation', *International Journal of Electronic Customer Relationship Management*, Vol. 1, No. 4, pp.341–353.

Barakat, N.H. and Bradley, A.P. (2007) 'Rule extraction from support vector machines: a sequential covering approach', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 19, No. 6, pp.729–741.

Barakat, N.H. and Bradley, A.P. (2010) 'Rule extraction from support vector machines: a review', *Neurocomputing*, Vol. 74, Nos. 1–3, pp.178–190.

Barakat, N.H. and Diederich, J. (2004) 'Learning-based rule-extraction from support vector machines', in *ICCTA'2004, Proceedings of the 14th International Conference on Computer Theory and Applications*, Alexandria, Egypt.

Barakath, N.H. and Diederich, J. (2005) 'Eclectic rule-extraction from support vector machines', *International journal of Computer Intelligence*, Vol. 2, No. 1, pp.59–62.

Bhatnagar, V. and Ranjan, J. (2010) 'Principles for successful mobile CRM in organisations from data mining perspective', *International Journal of Electronic Customer Relationship Management*, Vol. 4, No. 3, pp.280–301.

Bolton, R.N., Kannan, P.K. and Bramlett, M.D. (2005) 'Implications of loyalty program membership and service experiences for customer retention and value', *Journal of the Academy of Marketing Science*, Vol. 28, No. 1, pp.95–108.

Brockett, P., Derrig, R., Golden, L., Levine, A. and Alpert, M. (2002) 'Fraud classification using principal component analysis of RIDITs', *Journal of Risk and Insurance*, Vol. 69, No. 3, pp.341–371.

Bull, C. (2010) 'Customer relationship management systems, intermediation and disintermediation: the case of INSG', *International Journal of Information Management*, Vol. 30, No. 1, pp.94–97.

Business Intelligence Cup (2004) Organized by the University of Chile, available at http://www.tis.cl/bicup_04/text-bicup/BICUP/202004/20public/20data.zip (accessed on July 2007).

Chan, S.L. and Ip, W.H. (2011) 'A dynamic decision support system to predict the value of customer for new product development', *Decision Support System*, Vol. 52, No. 1, pp.178–188.

Chaves, Ad.C.F., Vellasco, M.M.B.R. and Tanscheit, R. (2005) 'Fuzzy rule extraction from support vector machines', in *ICHIS'2005, Proceedings of the Fifth International Conference on Hybrid Intelligent Systems*, Rio de Janeiro, Brazil.

Chen, Z., Li, J. and Wei, L. (2007) 'A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue', *Artificial Intelligence in Medicine*, Vol. 41, No. 2, pp.161–175.

Faase, R., Helms, R. and Spruit, M. (2011) 'Web 2.0 in the CRM domain: defining social CRM', *International Journal of Electronic Customer Relationship Management*, Vol. 5, No. 1, pp.1–22.

Farquad, M.A.H., Ravi, V. and Raju, S.B. (2008) 'Rule Extraction using support vector machine based hybrid classifier', in *TENCON-2008, IEEE Region 10 Conference*, Hyderabad, India.

Farquad, M.A.H., Ravi, V. and Raju, S.B. (2009) 'Data mining using rules extracted from SVM: an application to churn prediction in bank credit cards', in *RSFDGrC'2009, Proceedings of the 12th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, LNAI*, Delhi, India, Vol. 5908, pp.390–397.

Farquad, M.A.H., Ravi, V. and Raju, S.B. (2010a) '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, USA, Vol. 2, pp.404–426.

Farquad, M.A.H., Ravi, V. and Raju, S.B. (2010b) 'Rule extraction from support vector machine using modified active learning based approach: an application to CRM', in *KES'2010, Proceedings of the 14th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, Cardiff, Wales, UK.

Farquad, M.A.H., Ravi, V. and Raju, S.B. (2010c) 'Support vector regression based hybrid rule extraction methods for forecasting', *Expert Systems with Applications*, Vol. 37, No. 8, pp.5577–5589

Farquad, M.A.H., Ravi, V. and Raju, S.B. (2010d) 'Rule extraction from support vector machines: a hybrid approach for classification and regression problems', *International Journal of Information and Decision Sciences*, Vol. 3, No. 3, pp.265–294.

Farvaresh, H. and Sepehri, M.M. (2011) 'A data mining framework for detecting subscription fraud in telecommunication', *Engineering Applications of Artificial Intelligence*, Vol. 24, No. 1, pp.182–194.

Fawcett, T. (2006) 'An introduction to ROC analysis', *Pattern Recognition Letters*, Vol. 27, No. 8, pp.861–874.

Fawcett, T. and Provost, F. (1997) 'Adaptive fraud detection', *Data mining and Knowledge Discovery*, Vol. 1, No. 3, pp.291–316.

Fu, X., Ong, C.J., Keerthi, S., Hung, G.G. and Goh, L. (2004) 'Extracting the knowledge embedded in support vector machines', in *IJCNN'04, Proceedings of the International Joint Conference on Neural Networks*, Budapest, Hungary.

Fung, G., Sandilya, S. and Bharat, R.R. (2005) 'Rule extraction from linear support vector machines', in *SIGKDD'05, Proceedings of the Eleventh ACM International Conference on Knowledge Discovery in Data Mining*, ACM Press, New York, NY, USA, pp.32–40.

Gallant, S. (1988) 'Connectionist expert systems', *Communications of the ACM*, Vol. 31, No. 2, pp.152–169.

Hayashi, Y., Hsieh, M-H. and Setiono, R. (2010) 'Understanding consumer heterogeneity: a business intelligence application of neural networks', *Knowledge Based Systems*, Vol. 23, No. 8, pp.856–863.

Huang, B., Kechadi, M. and Buckley, B. (2012) 'Customer churn prediction in telecommunications', *Expert Systems with Applications*, Vol. 39, No. 1, pp.1414–1425.

Kohavi, R. (1996) 'Scaling up the accuracy of naïve-Bayes classifiers: a decision-tree hybrid', in *KDD'96, Proceedings of the International Conference on Knowledge Discovery and Data Mining*, Portland, USA.

Kotler, P. (2002) *Marketing Management: The Millennium Edition*, Prentice-Hall International, Englewood Cliffs, NJ.

Kumar, D.A. and Ravi, V. (2008) 'Predicting credit card customer churn in banks using data mining', *International Journal for Data Analysis, Techniques and Strategies*, Vol. 1, No. 1, pp.4–28.

Larivière, B. and Van den Poel, D. (2004) 'Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: the case of financial services', *Expert Systems with Applications*, Vol. 27, No. 2, pp.277–285.

Martens, D., Baesens, B. and Gestel, T.V. (2009) 'Decompositional rule extraction from support vector machines by active learning', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, No. 2, pp.178–191.

Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M. and Euler, T. (2006) 'YALE: rapid prototyping for complex data mining tasks', in *KDD'06, Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, USA.

Mozer, M.C., Wolniewicz, R., Grimes, D.B., Johnson, E. and Kaushansky, H. (2000) 'Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry', *IEEE Transactions of Neural Networks*, Vol. 11, No. 3, pp.690–696.

Muther, A. (2002) *Customer Relationship Management: Electronic Customer Care in the New Economy*, Spinger-Verlag, Berlin.

Naveen, N., Ravi, V. and Kumar, D.A. (2009) 'Application of fuzzy ARTMAP for churn prediction in bank credit cards', *International Journal of Information and Decision Sciences*, Vol. 1, No. 4, pp.428–444.

Nie, G., Rowe, W., Zhang, L., Tian, Y. and Shi, Y. (2011) 'Credit card churn forecasting by logistic regression and decision tree', *Expert System with Applications*, Vol. 38, No. 12, pp.15273–15285.

Nunez, H., Angulo, C. and Catata, A. (2002) 'Rule extraction from support vector machines', in *ESANN'02, Proceedings of the European Symposium on Artificial Neural Networks*, pp.107–112.

Ojasalo, J. (2002) 'Key account management in information-intensive services', *Journal of Retailing and Consumer Services*, Vol. 9, No. 5, pp.269–276.

Phua, C., Damminda, A. and Lee, V. (2004) 'Minority report in fraud detection: classification of skewed data', in *SIGKDD'04, Explorations: Special Issue on Imbalanced Data Sets*, Vol. 6, No. 1, pp.50–59.

Phua, C., Lee, V., Smith, K. and Gayler, R. (2010) 'A comprehensive survey of data mining-based fraud detection research', *Artificial Intelligence Review*, available at http://arxiv.org/abs/1009.6119 (accessed on 15 July 2011).

Pyle, D. (1999) *Data Preparation for Data Mining*, Morgan Kaufmann Publishers, San Francisco, USA.

Quinlan, J.R. (1986) 'Induction of decision trees', *Machine Learning*, Vol. 1, No. 1, pp.81–106.

Shaw, M., Subramaniam, C., Tan, G.W. and Welge, M. (2001) 'Knowledge management and data mining for marketing', *Decision Support Systems*, Vol. 31, No. 1, pp.127–137.

Tickle, A.B., Andrews, R., Golea, M. and Diederich, J. (1998) 'The truth will come to light: directions and challenges in extracting the knowledge embedded within trained artificial neural network', *IEEE Transactions on Neural Networks*, Vol. 9, No. 6, pp.1057–1068.

Vapnik, V.N. (1995) *The Nature of Statistical Learning Theory*, Springer-Verlag, New York Inc., NY, USA.

Viaene, S., Derrig, R. and Dedene, G. (2004) 'A case study of applying boosting Naive Bayes to claim fraud diagnosis', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No. 5, pp.612–620.

Wheeler, R. and Aitken, S. (2000) 'Multiple algorithms for fraud detection', *Knowledge-Based Systems*, Vol. 13, Nos. 2–3, pp.93–99.

Wong, W.K., Leung, S.Y.S., Guo, Z.X., Zeng, X.H. and Mok, P.Y. (2012) 'Intelligent product cross-selling system with radio frequency identification technology for retailing', *International Journal of production Economics*, Vol. 135, No. 1, pp.308–319.

Xia, G-E. and Jin, W-D. (2008) 'Model of customer churn prediction on support vector machine', *System Engineering – Theory & Practice*, Vol. 28, No. 1, pp.71–77.

Yu, X., Guo, S., Guo, J. and Huang, X. (2011) 'An extended support vector machine forecasting framework for customer churn in e-commerce', *Expert Systems with Applications*, Vol. 38, No. 3, pp.1425–1430.

Zhang, Y., Su, H., Jia, T. and Chu, J. (2005) 'Rule extraction from trained support vector machines', *Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, Vol. 3518, pp.61–70.