An Application of Support Vector Machines for Customer Churn Analysis: Credit Card Case

Sun Kim, Kyung-shik Shin, and Kyungdo Park

Ewha Womans University, College of Business Administration,
11-1 Daehyun-Dong, Seodaemun-Gu, Seoul 120-750, Korea
kimsun0122@empal.com, {ksshin, kyungdo}@ewha.ac.kr

Abstract. This study investigates the effectiveness of support vector machines (SVM) approach in detecting the underlying data pattern for the credit card customer churn analysis. This article introduces a relatively new machine learning technique, SVM, to the customer churning problem in attempt to provide a model with better prediction accuracy. To compare the performance of the proposed model, we used a widely adopted and applied Artificial Intelligence (AI) method, back-propagation neural networks (BPN) as a benchmark. The results demonstrate that SVM outperforms BPN. We also examine the effect of the variability in performance with respect to various values of parameters in SVM.

1 Introduction

Increasing the customer retention rate using the customer databases is one of the major concerns among marketing managers. It is widely accepted that the cost of retaining current customers are much cheaper than the cost of obtaining new customers. Due to the high cost of acquiring new customers and considerable benefits of retaining existing ones, building a churn prediction model to facilitate subsequent churn management and customer retention is critical for the success of the firms facing competitive market environment. The economic value of customer retention has been demonstrated in several empirical research applied to financial industry. For example, Reichheld and Sasser (1990) [19] found that a bank is able to increase its profits by 85% due to a 5 % improvement in the retention rate. Similar findings were obtained in Van den Poel and Lariviere (2004) [14], where the financial impacts of one percent increase in customer retention rate were calculated.

This study investigates the effectiveness of support vector machines (SVM) approach in detecting the underlying data pattern for the credit card customer churn analysis. SVM classification exercise finds hyperplanes in the possible space for maximizing the distance from the hyperplane to the data points, which is equivalent to solving a quadratic optimization problem. The solution of strictly convex problems for support vector machines is unique and global. SVM implements the structural risk minimization (SRM) principle that is known to have high generalization performance. As the complexity increases by numbers of support vectors, SVM is constructed through trading off decreasing the number of training errors and increasing the risk of over-fitting the data.

L. Wang, K. Chen, and Y.S. Ong (Eds.): ICNC 2005, LNCS 3611, pp. 636 – 647, 2005.

Since SVM captures geometric characteristics of feature space without deriving weights of networks from the training data, it is capable of extracting the optimal solution with the small training set size.

While there are several arguments that support the observed high accuracy of SVM, the preliminary results show that the accuracy and generalization performance of SVM is better than that of the standard back-propagation neural networks. In addition, since choosing an appropriate value for parameters of SVM plays an important role on the performance of SVM, we also investigate the effect of the variability in prediction and generalization performance of SVM with respect to various values of parameters in SVM such as the upper bound C and the bandwidth of the kernel function.

The remainder of this paper is organized as follows. Related studies about credit card research and the business application using support vector machines are provided in section 2. Section 3 provides a brief description of the research methods. In this section we also demonstrate the several superior points of the SVM algorithm compared with BPN. Section 4 describes the research data and experiments. Section 5 summarizes and analyzes empirical results. Section 6 discusses the conclusions and future research issues.

2 Related Studies

Customer churn prediction and management is a concern for many industries, but it is particularly acute in the strongly competitive and now broadly liberalized mobile telecommunications industry. A mobile service provider wishing to retain it's subscribers needs to be able to predict which of them may be at-risk of changing services and will make those subscribers the focus of customer retention efforts [28].

On the contrary, there are relatively few studies on the customer churn analysis of credit card holders. Lee *et al*. (2001) [15] proposed a fuzzy cognitive map approach to integrate explicit knowledge and tacit knowledge for churn analysis of credit card holders in Korea. Lee *et al*. (2002) [16] compared the neural network approach with logistic analysis, and C5.0 for churn analysis of credit card holders in Korea.

SVM has shown excellent generalization performance on a wide range of problems including bioinformatics [2] [11] [30], text categorization [12], face detection using image [18], hand written digit recognition [4] [5], medical diagnosis [22], estimating manufacturing yields [21]. These application domains typically involved high-dimensional input space, and the good performance is also related to the fact that SVM's learning ability can be independent of the dimensionality of the feature pace [10].

The SVM approach has been several business applications recently, mainly in the area of time series prediction and classification [9] [13] [17] [23] [24] [25], marketing [1], bankruptcy prediction [20], credit rating analysis [10]. However, there is no research using SVM to credit card customer data. This study is the first attempt of using SVM to credit card customer databases.

3 Support Vector Machines

SVM is a new learning machine method introduced first by Vapnik. The basic SVM deals with two-class problems in which the data are separated by a hyperplane defined by a number of support vectors. It is based on the Structural Risk Minimization (SRM) principle from computational learning theory [26] [27].

The underlying theme of the class of supervised learning method is to learn from observations. SVM produces a binary classifier, the so-called optimal separating hyperplanes, through nonlinear mapping of the input vectors into the high-dimensional feature space. SVM constructs linear model to estimate the decision function using nonlinear class boundaries based on support vectors. If the data is linearly separated, SVM trains linear machines for an optimal hyperplane that separates the data without error and into the maximum distance between the hyperplane and the closest training points. The training points that are closest to the optimal separating hyperplane are called support vectors. All other training examples are irrelevant for determining the binary class boundaries. In general cases where the data is not linearly separated, SVM uses nonlinear machines to find a hyperplane that minimize the number of errors for the training set [20].

Let the labeled training examples as $[x_i, y_i]$, an input vector as $x_i \in R^n$, and the class value as $y_i \in \{-1, 1\}$, for i =1,…l. For the linearly separable case, the decision rules defined by an optimal hyperplane separating the binary decision classis is given as:

$$Y = sign\ (\sum_{i=1}^{N} y_i a_i (x \cdot x_i) + b) \cdot \tag{1}$$

where Y is the outcome, y_t is the class value of the training example x_i, which represents the inner product. The vector $x = (x_1, x_2, \ldots\ldots, x_n)$ corresponds to an input and the vectors x_i, i=1…N, are the support vectors. In the Equation (1), b and a_i are parameters that determine the hyperplane.

For the nonlinearly separable case, a high-dimensional version of Equation (1) is given as follows:

$$Y = sign\ (\sum_{i=1}^{N} y_i a_i K\ (x, x_i) + b) \cdot \tag{2}$$

In equation (2), the function $K(x, x_i)$ is defined as the kernel function for generating the inner products to construct machines with different types of nonlinear decision surfaces in the input space. For construction the decision rules, three common types of SVM are given as Table 1.

Table 1. Commonly used Kernel functions

Name	Mathematical form*
Polynomial kernel	$K(x,x_i) = (x \cdot x_i + 1)^d$
Radial basis function kernel	$K(x,x_i) = \exp(-\frac{1}{\delta^2 (x - x_i)^2})$
Two-layer neural network kernel	$K(x,x_i) = S[(x \cdot x_i)] = \frac{1}{[1 + \exp\{v(x \cdot x_i) - c\}]}$

* d is the degree of the polynomial kernel.
δ^2 is the bandwidth of the radial basis function kernel.
v and c are parameters of a sigmoid function $S[(x \cdot x_i)]$ satisfying the inequality $c \geq v$.

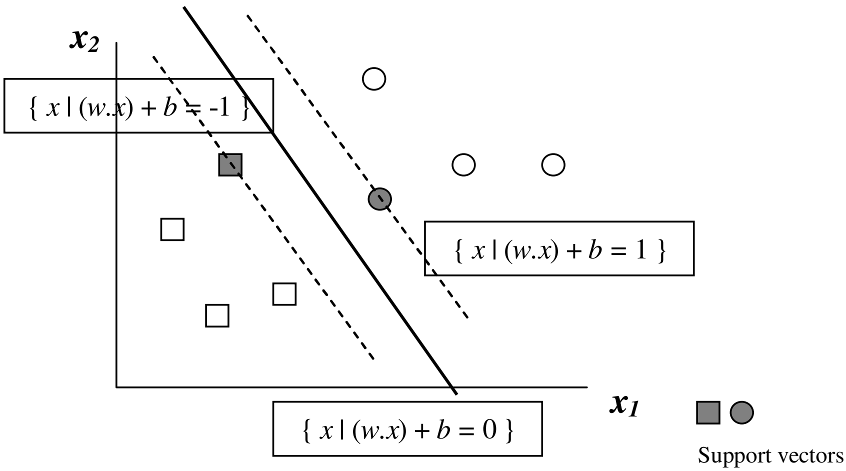

Fig. 1. Classification of data by SVM – Linear case

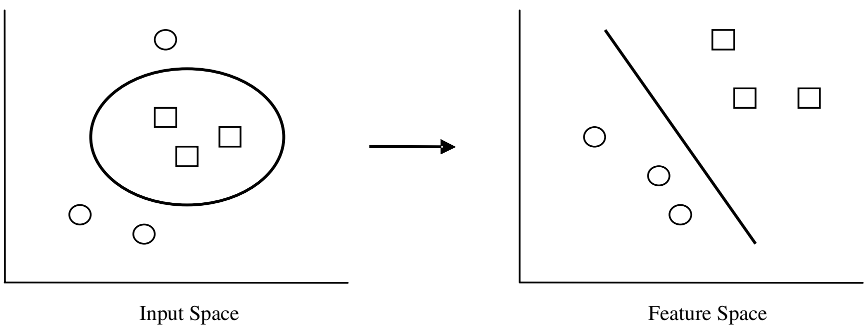

Fig. 2. Non-linear separation of input and feature space

The SVM classification exercise is implemented in solving a linearly constrained quadratic programming (QP) for finding the support vectors and determining the parameters b and a_i. For the separable case, there is a lower bound 0 on the coefficient a_i in Equation (1). For the non-separable case, SVM can be generalized by placing an upper bound C on the coefficients a_i. in addition to the lower bound [29].

In brief, the learning process to construct decision functions of SVM is completely represented by the structure of two layers, which seems to be similar with BPN. However, learning algorithm is different in that SVM is trained with optimization theory that minimizes misclassification based on statistical learning theory. The first layer selects the basis $K(x, x_i)$, $i = 1 \ldots N$ and the number of support vectors from given set of bases defined be the kernel. The second layer constructs the optimal hyperplane in the corresponding feature space [27]. The scheme of SVM is shown in Figure 3.

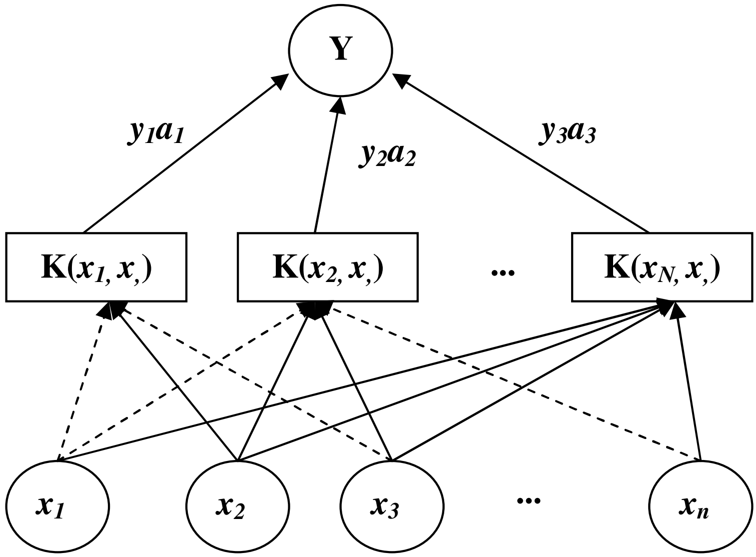

Fig. 3. The scheme of SVM (adapted from [26])

Compared with the limitations of the BPN, the major advantages of SVM are as follows. First, SVM has only two free parameters, namely the upper bound and kernel parameter. On the other hand, because a large number of controlling parameters in BPN such as the number of hidden layers, the number of hidden nodes, the learning rate, the momentum term, epochs, transfer functions and weights initialization methods are selected empirically. Therefore, in BPN, it is a difficult task to obtain an optimal combination of parameters that produces the best prediction performance.

Second, SVM guarantees the existence of unique, optimal and global solution since the training of SVM is equivalent to solving a linearly constrained QP. On the other hand, because the gradient descent algorithm optimizes the weights of BPN in a way that the sum of square error is minimized along the steepest slope the error surface, the result from training may be massively multi modal, leading to non-unique solutions, and be in the danger of getting stuck in a local minima.

Third, SVM implement the structural risk minimization (SRM) principle that is known to have a good generalization performance. SRM is the approach to trading off empirical error with the capacity of the set called VC dimension, which seeks to minimize an upper bound of the generalization error rather than minimize the training error. In order to apply SRM, the hierarchy of hypothesis spaces must be defined before the data is observed. But in SVM, the data is first used to decide which hierarchy to use and then subsequently to find a best hypothesis from each. Therefore the argument that good generalization performance of SVM is attributable to SRM is flawed, since the result of SVM is obtained from a data dependent SRM [3].

Although the other reason why SVM has good generalization performance is suggested [27], there exists no explicitly established theory that shows good generalization performance is guaranteed for SVM. However, it seems plausible that performance of SVM is more general than that of BPN because the two measures in terms of the margin and number of support vectors give information about the relation between the input and target function according to different criteria, either of which is sufficient to indicate good generalization. On the other hand, BPN is based on minimizing a squared error criterion at the network output, and tends to produce a classifier with the only large margin measure. In addition, flexibility caused by choosing training data is likely to occur with weights of BPN model, but the maximum hyperplane of SVM is relatively stable and gives little flexibility in the decision boundary [29].

Finally, SVM is constructed with the small training data set size, since it learns by capturing geometric picture corresponding to the kernel function. Moreover, no matter how large the training size is, SVM is capable of extracting the optimal solution with the small training set size. On the other hand, for the case of BPN containing a single hidden layer and used as a binary classifier, it is provided that the number of training examples, with an error of 10 percent, should be approximately 10 times the number of weights in the network. With 10 input and hidden nodes, the learning algorithm will need more than 1,000 training set size that is sufficient for a good generalization [7]. However, in most practical applications, there can be a huge numerical gap between the actual size of the training set needed and that is available.

Due to utilizing the feature space images by the kernel function, SVM is applicable in such circumstances that have proved difficult or impossible for BPN where data in the plane is randomly scattered and the density of the data's distribution is not even well defined [6].

4 Research Data and Experiments

4.1 Data Description and Variable Selection Procedure

For the purpose of this study, we prepare the data from the credit card company in Korea. We obtain the data about the customers who retain their credit card from April 1997 to October 2000 and the customers who close their account during the same period. The data set covers demographic variables and the variables about credit card usage. After filtering the data with missing values, we select 4,650 samples for each case. The description of the variables for this research is presented in Table 2.

Table 2. Definition of variables

Variable	Definition
x_1 MON2REN	Month to Renewal
x_2 AVG_CLS	Average Line Size (Credit Limit)
x_3 AGE	Age in Years
x_4 GENDER	Male or Female
x_5 AVERAGE	Average Usage Amount
x_6 AGEING	Installment Period
x_7 INTEREST	Average Interest
y STATUS	Holding or not

In this study we limit usage amount and arrear data to 3 months data because we believe a 3-month period would be sufficient time to understand customers' credit status and behavior. That is, credit card companies register a customer that arrears the payment over 3 months into credit defaulter. In this reason the usage amount data we employ is the 3 months average usage amount. We also use the demographic variables such as age and gender (male:0, female:1), categorizing ageing data into 6 installment periods (0, 3, 6, 9, 12, 15). Two status (closing their account: 0, retaining credit card: 1) appear in our data set as a dependent variable. The total sample size is 9,210.

Table 3. The number of sample data

	The Number of Sample Data
Customers that retained credit card	4,605
Customers that closed their account	4,605
Total	9,210

4.2 Experiments

The data set is arbitrarily split into two subsets; about 80% of the data is used for a training set and 20% for a validation set. The training data for SVM is entirely used to construct the model while BPN is divided into 60% training set and 20% test set. We prepare five data sets to conduct the experiment as table 4.

Table 4. Training set and validation set

	SVM		BPN	
	# of Sample	Ratio	# of Sample	Ratio
Training set	800	80%	600	60%
Test set			200	20%
Validation set	200	20%	200	20%
Total	1,000	100%	1,000	100%

In this study, the radial basis function is used as the kernel function of SVM. Since SVM does not have a general guidance for determining the upper bound C and kernel parameter $\gamma\left(=\frac{1}{\delta^2}\right)$, this study varies the parameters to select optimal values for the best prediction performance. We use the software package Hsu and Lin (2001) [8] provided, BSVM, for our study.

To verify the applicability of SVM, we use BPN as the benchmark with the following controlling parameters. The structure of BPN is a standard three-layer with the same number of input nodes in the hidden layer and the hidden and output nodes use the sigmoid transfer function. For stopping the training of BPN, test set that is not a subset of the training set is used. However the optimum network for the data in the test set is still different to guarantee generalization performance. The neural network algorithms software NeuroShell 2 version 4.0 executes these processes.

5 Results and Analysis

To investigate the effectiveness of the SVM approach on the churn analysis, we conduct the experiment with respect to various kernel parameters and the upper bound C, and compare the prediction performance of SVM with various parameters. Based on the results proposed by Tay and Cao (2001) [23] and Shin, Lee, and Kim (2005) [20], we set an appropriate range of parameters as follows: a range for kernel parameter $\gamma\,(=\frac{1}{\delta^2})$ is between 0.1 and 0.008 and a range for C is between 1 and100. The results are summarized in Table 5. Each cell of Table 5 contains the accuracy of the classification techniques.

Table 5. Classification accuracies (%) of various parameters in SVM

C	γ=0.1		γ=0.05		γ=0.02		γ=0.01		γ=0.008	
	Tr.	Val.	Tr.	Val.	Tr.	Val.	Tr.	Val.	Tr.	Val.
1st										
1	94.3	82.0	94.2	82.0	94.1	82.0	93.2	79.0	93.2	79.5
10	92.2	82.5	95.0	82.0	94.3	80.5	94.5	81.0	94.3	80.5
50	95.7	83.0	95.2	82.5	94.5	79.5	94.5	79.5	94.5	79.5
75	96.1	81.0	95.5	82.5	95.0	81.0	94.5	79.5	94.5	79.0
100	96.1	81.0	95.5	83.0	95.0	81.0	94.5	79.5	94.5	79.0
2nd										
1	95.7	74.0	94.7	76.0	94.5	77.5	94.8	78.5	94.8	78.5
10	96.1	76.5	96.1	77.0	95.8	77.5	95.5	78.0	95.0	78.0
50	96.6	75.5	96.2	77.0	96.1	77.5	96.0	78.0	95.5	78.0
75	96.7	75.5	96.2	76.5	96.0	77.5	96.0	78.0	96.0	78.0
100	96.7	76.0	96.3	76.0	96.0	77.5	96.0	78.0	96.1	77.5
3rd										
1	94.7	83.5	94.6	83.0	95.5	80.5	94.2	83.5	94.0	83.0
10	95.7	81.5	95.0	84.0	95.0	81.0	94.4	84.5	94.3	84.0
50	96.3	79.5	95.7	82.0	95.1	80.5	95.0	84.0	94.5	84.5
75	96.5	79.5	95.7	82.0	95.0	80.0	94.8	83.5	94.5	84.5
100	96.6	79.5	95.7	82.0	95.0	80.0	94.8	83.5	94.8	84.0
4th										
1	95.1	79.0	94.3	80.0	94.0	80.5	93.5	81.5	93.5	81.5
10	95.8	78.0	95.2	80.0	94.8	81.0	94.2	81.5	94.3	81.5
50	96.2	77.5	96.0	79.5	95.5	80.5	94.8	81.0	94.7	81.5
75	96.5	76.5	96.1	78.5	95.5	80.0	94.8	81.5	94.8	81.5
100	96.6	75.5	96.1	78.5	95.5	80.0	95.1	81.5	94.7	81.0
5th										
1	95.5	79.0	95.0	79.0	94.7	80.0	94.6	80.0	94.5	81.0
10	96.5	78.0	96.2	79.0	95.7	79.5	94.8	83.5	94.8	83.5
50	97.2	76.5	96.8	78.0	95.8	80.0	95.5	84.5	95.5	84.5
75	97.0	76.0	96.7	78.5	95.8	80.0	95.5	84.5	95.5	84.5
100	97.2	76.5	96.6	78.5	96.1	78.5	95.7	84.0	95.5	84.5

The experimental results show that the overall prediction performance of SVM is sensitive not only to various data sets but also to various parameters such as the kernel parameter γ and the upper bound C. In Table 5, the results of SVM on the validation set show the best prediction performances when γ is 0.01 on the most data set except 1st set.

The accuracy on the training set increases monotonically as C increases; on the contrary, the accuracy on the validation set shows a tendency to increase slightly. This indicates that a large value for C has an inclination to over-fit the training data and an appropriate value for C plays a leading role on preventing SVM from deterioration in the generalization performance [20]. According to Tay and Cao (2001) [23] , a small value for C would under-fit the training data because the weight placed on the training data is too small and leads to small values of prediction accuracy on both the training and validation sets while a large value for C would over-fit the training data. In this study, the prediction performance on the training set increases as C increases while the prediction performance on the validation set maintains an almost constant value as C increases. These results partly support the conclusion of Tay and Cao (2001) [23]. Figure 2 gives the results of SVM on the 2nd data set with various C where γ is fixed at 0.008.

Figure 6 gives the results of SVM on the 5th data set with various γ where C is fixed at 75. The accuracy on the training set of the most data set except 1st set decreases as γ decreases; on the other hand, the accuracy on the validation set shows a tendency to increase with decreasing γ. According to Shin, Lee, and Kim (2005) [20], this indicates that a large value for γ has an inclination to over-fit the training data and an appropriate value for γ also plays an important role on the generalization performance of SVM. These results also support the conclusion of Tay and Cao (2001) [23].

Another focus of this study is on the comparison of prediction accuracies between SVM and BPN. The results of the best SVM model that present the best prediction performance for the validation set from five data sets and average of all sets are compared with those of BPN and are summarized in Table 6. Each cell of the table contains the accuracy of the classification techniques.

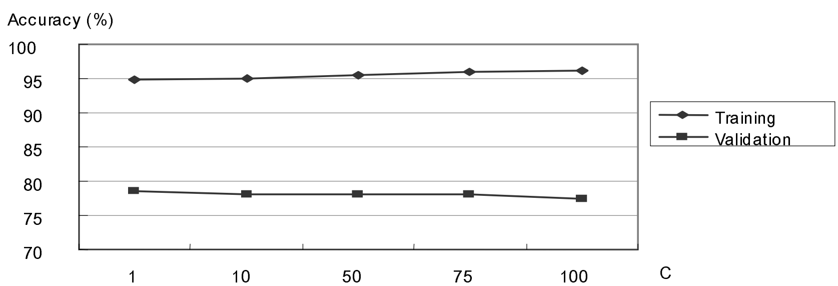

Fig. 5. Results of SVM with various C where γ is fixed at 0.008 on the 2nd set

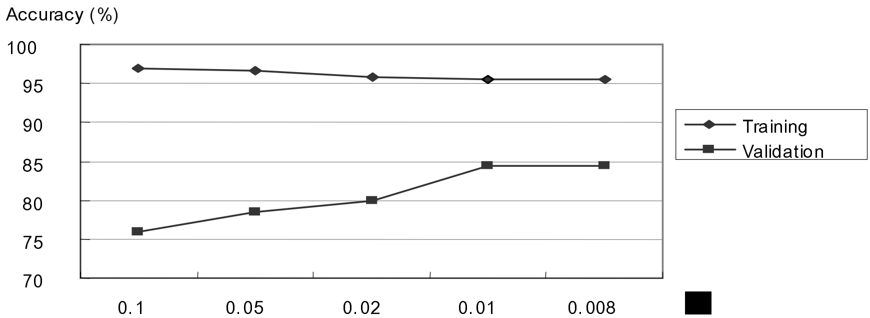

Fig. 6. Results of SVM with various γ where C is fixed at 75 on the 5th set

In Table 6, SVM slightly improved the churning prediction accuracies on all of the data sets and the average. Using SVM, the accuracy (%) on the training set is over 90% and that on the validation set is around 82%. On the other hand, when we use BPN, the accuracy (%) on both of the training and validation set are around 79%. The results show that SVM outperforms BPN in terms of prediction rates

Table 6. Comparison of classification accuracies between the best SVM and BPN

Data set		SVM		BPN	
		Accuracy	%	Accuracy	%
1st set	Training	764	95.5	472	78.7
	Test			159	79.5
	Validation	166	83.0	160	80.0
2nd set	Training	758	94.8	475	79.2
	Test			157	78.5
	Validation	157	78.5	150	75.0
3rd set	Training	755	94.4	484	80.7
	Test			165	82.5
	Validation	169	84.5	163	81.5
4th set	Training	748	93.5	479	79.8
	Test			162	81.0
	Validation	163	81.5	161	80.5
5th set	Training	764	95.5	476	79.3
	Test			161	80.5
	Validation	169	84.5	158	79.0
Total set	Training	3,789	94.7	2,386	79.5
(Avg. %)	Test			804	80.4
	Validation	824	82.4	792	79.2

6 Conclusions

In this study, we applied a newly introduced learning method, support vector machines (SVM), together with a frequently used high performance method, back-propagation neural networks (BPN), to the problem of he credit card customer churn analysis. Although it is the preliminary research, we can draw several conclusions from this experiment. Our results demonstrate that SVM has the higher level of prediction performance than BPN. We also examine the effect of various values of parameters in SVM such as the upper bound C and the bandwidth of the kernel function. We found that the prediction performance is sensitive to the values of these parameters. This result suggests that it is important to optimize the kernel function and various parameters simultaneously because the determination of all these parameter values has critical impacts on the performance of the resulting system. Therefore, developing the structured method of selecting the optimal parameter values for SVM is necessary to obtain the best prediction performance.

References

1. Ben-David, S. and Lindenbaum, M.: Learning distributions by their density levels: A paradigm for learning without a teacher, Vol. 55. Journal of Computer and System Sciences (1997) 171-182
2. Brown, M.P., Grudy W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., Haussler, D.: Knowledge-based analysis of microarray gene expression data by using support vector machines, Vol. 97(1). Proceedings of National Academy of Sciences (2000) 262-267
3. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition, Vol. 2(2). Data Mining and Knowledge Discovery (1998) 955-974
4. Burges, C.J.C., and Scholkopf, B.: Improving the accuracy and speed of support vector machines. In Mozer, M., Jordan, M., and Petche, T., Advances in Neural information processing systems. Cambridge, MA, MIT Press (1997)
5. Cortes, C., and Vapnik, V.N.: Support vector networks, Vol. 20. machine Learning (1995) 273-297
6. Friedman, C.: Credit model technical white paper. Standard and Poor's. New York: McGraw-Hill. (2002)
7. Haykin, S.: Neural networks: A comprehensive foundation. New York: Macmillan (1994)
8. Hsu, C.W. and Lin, C.J.: A comparison of methods for multi-class support vector machines, Technical Report, National Taiwan University, Taiwan, 2001 (2001)
9. Huang, W., Nakamori, Y., Wang, S.Y.: Forecasting stock market movement direction with support vector machine, Computers & Operations Research (2004)
10. Huang, Z., Chen, H., Hsu, C.J., Chen, W.H., Wu, S.: Credit rating analysis with support vector machines and neural networks: a market comparative study, Vol. 37. Decision Support Systems (2004) 543-558
11. Jaakkola, T.S. and Haussler, D.: Esploiting generative models in discriminative classifiers, in: M.S. Kearns, S.A. Solla, D.A. Cohn (Eds.), Advances in Neural Information Processing Systems, MIT Press, Cambridge (1998)
12. Joachims, T.: Learning to classify text using support vector machines, London: Kluwer Academic Publishers (2002)

13. Kim, K.J.: Financial time series forecasting using support vector machines, Vol. 55(1-2). Neurocomputing (2003) 307-319
14. Lariviere, B. and Van den Poel, D.: Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: the case of financial services, Vol. 27. Expert Systems with Applications (2004) 277-285
15. Lee, K.C., Chung, N.H., Kim, J.K.: A fuzzy cognitive map approach to integrating explicit knowledge and tacit knowledge: emphasis on the churn analysis of credit card holders, Vol. 11(4). Information Systems Review (2001) 113-133
16. Lee, K.C., Chung, N.H., Shin, K.S.: An artificial intelligence-based data mining approach to extracting strategies for reducing the churning rate in credit card industry, Vol. 8(2). Journal of Intelligent Information Systems (2002) 15-35
17. Mukherjee, S., Osuna, E., and Girosi, F.: Nonlinear prediction of chaotic time series using support vector, Proceedings of the IEEE Workshop on Neural Networks for Signal Processing (1997) 511-520
18. Osuna, E. Freund, R., and Girosi, F.: Training support vector machines: An application to face detection, Proceedings of Computer Vision and Pattern Recognition (1997) 130-136
19. Reichheld, F.F., & Sasser, W.E., Jr.: Zero defections: quality comes to service., Vol. 68(5). Harvard Business Review (1990) 105-111
20. Shin, K.S., Lee, T.S., and Kim, H.J.: An application of support vector machines in bankruptcy prediction model, Vol. 28. Expert Systems with Applications (2005) 127-135
21. Stoneking, D.: Improving the manufacturability of electronic designs, Vol. 36(6). IEEE Spectrum (1999) 70-76
22. Tarassenko, L., Hayton, P., Cerneaz, N., and Brady, M.: Novelty detection for the identification of masses in mammograms, Proceedings Fourth IEEE International Conference on Arificial Neural Networks, Cambridge (1995) 442-447
23. Tay, F.E.H. and Cao, L.: Application of support vector machines in financial time series forecasting, Vol. 29. Omega (2001) 309-317
24. Tay, F.E.H. and Cao, L.J.: Modified support vector machines in financial time series forecasting, Vol. 48. Nerocomputing (2002) 847-861
25. Van Gestel, T. Suykens, J.A.K., Baestaens, D.-E., Lambrachts, A., Lanckriet, G., Vandaele, B. De Moor, B., Vandewalle, J.: Financial time series prediction using least squares support vector machines within the evidence frame work, Vol. 12(4). IEEE Transactions on Neural Networks (2001) 809-821
26. Vapnik, V.N.: The nature of Statistical learning Theory. New York: Springer-Verlag. (1995)
27. Vapnik, V.N.: Statistical learning theory. New York: John Wiley and Sons (1998)
28. Wei, C.P and Chiu, I.T.: Turning telecommunications call details to churn prediction: a data mining approach, Vol. 23. Expert Systems with Applications (2002) 103-112
29. Witten, I.H., and Frank, E.: Data Mining: Practical machine learning tools and techniques with java implementations. San Francisco, CA: Morgan Kaufmann Publishers. (2000)
30. Zien, A., Ratsch, G., Mika, S., Scholkopf, B., Lengauer,T., Muller, K.-R.: Engineering support vector machine kernels that recognize translation initiation sites, Vol. 16(9). Bioinformatics (2000) 799-807