Two dimensional reduction technique of Support Vector Machines for Bankruptcy Prediction

Hyunchul Ahn[a], Kyoung-jae Kim[b] and Kichun Lee[c]

[a] *Center for Military Planning, Korea Institute for Defense Analyses*
5-7, Cheongrangri 2-Dong, Dongdaemun-Gu, Seoul, 130-012, Korea
Tel: +82-2-961-1335, Fax: +82-2-961-1163, E-mail: hcahn@kida.re.kr

[b] *Department of Management Information Systems, Dongguk University*
3-26, Pil-Dong, Chung-Gu, Seoul, 100-715, Korea
Tel: +82-2-2260-3324, Fax: +82-2-2260-8824, E-mail: kjkim@dongguk.edu

[c] *H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology*
765 Ferst Drive, NW, Atlanta, Georgia 30332-0205, USA
Tel: +1-404-206-3554, Fax: +1-404-894-4243, E-mail: klee7@gatech.edu

Abstract

Prediction of corporate bankruptcies has long been an important topic and has been studied extensively in the finance and management literature because it is an essential basis for the risk management of financial institutions. Recently, support vector machines (SVMs) are becoming popular as a tool for bankruptcy prediction because they use a risk function consisting of the empirical error and a regularized term which is derived from the structural risk minimization principle. In addition, they don't require huge training samples and have little possibility of overfitting. However, in order to use SVM, a user should determine several factors such as the parameters of a kernel function, appropriate feature subset, and proper instance subset by heuristics, which hinders accurate prediction results when using SVM.
In this study, we propose a novel hybrid SVM classifier with simultaneous optimization of feature subsets, instance subsets, and kernel parameters. This study introduces genetic algorithms (GAs) to optimize the feature selection, instance selection, and kernel parameters simultaneously. Our study applies the proposed model to the real-world case for bankruptcy prediction. Experimental results show that the prediction accuracy of conventional SVM may be improved significantly by using our model.

Keywords:

Support Vector Machines; Genetic Algorithms; Bankruptcy Prediction

Introduction

Prediction of corporate bankruptcies has long been an important topic and has been studied extensively in the finance and management literature because it is an essential basis for the risk management of financial institutions. Bankruptcy prediction models have used various statistical and artificial intelligence techniques. These techniques include discriminant analysis, logistic regression, the decision tree, *k*-nearest neighbor, and backpropagation (BP) neural network. Among them, the BP network (BPN) has become one of the most popular techniques for the prediction of corporate bankruptcy due to its high prediction accuracy. However, many financial companies still have difficulties in using BPN. The difficulty stems from inherent limitations of BPN such as the requirement of large data samples, the possibility of overfitting, and poor explanatory power for the results.

Support vector machines (SVMs) may be an alternative to relieve these limitations of BPN [12]. General BPN models implement the empirical risk minimization principle for seeking to minimize the misclassification error or deviation from the correct solution of the training data. However, SVM implements the structural risk minimization principle for searching to minimize an upper bound of generalization error. In addition, the solution of SVM may be the global optimum, while BPN models may tend to fall into a local optimal solution. Therefore, overfitting of the results is unlikely to occur with SVM. Consequently, several recent studies for bankruptcy prediction used SVM as a classifier, and they showed that it might be an effective technique for predicting corporate financial distress [5,17,18,25].

However, SVM also has some factors that affect the prediction performance – these factors are usually set by heuristics. In particular, the selection of an appropriate kernel function and its parameters (e.g. C, d, δ^2) and the selection of proper feature subset in SVM have been popular research topics. Other than these factors, the selection of appropriate instance selection (in other words, prototype selection) may also improve the classification accuracy of SVM by eliminating irrelevant and distorting training samples. Nonetheless, there have been few studies that have applied instance selection to SVM, especially in the domain of bankruptcy prediction.

Thus, in this study, we propose a novel hybrid SVM classifier with simultaneous optimization of feature subsets, instance subsets, and kernel parameters. This study introduces genetic algorithms (GAs) to optimize the feature selection, instance selection, and kernel parameters simultaneously. Our study applies the proposed model to the real-world case for bankruptcy prediction, and presents experimental results from the application.

Literature Review

In this study, we propose the combined model of two artificial intelligence techniques, SVM and GA for effective bankruptcy prediction. Thus, in this section, we first review the prior studies on bankruptcy prediction and examine their limitations. After that, we review the basic concepts of SVM and GA, which are the core algorithms of our model. Finally, we introduce prior studies that attempt to optimize SVM using GA.

Prior Studies on Bankruptcy Prediction

There has been substantial research into bankruptcy prediction because it is one of the most important problems for companies and financial institutions. Various techniques including ANN, the decision tree, logistic regression (LOGIT), and discriminant analysis (DA) have been employed to predict corporate bankruptcy.
Early studies by Altman [1] used discriminant analysis to predict corporate bankruptcies. More recent research by Ohlson [21] used LOGIT and PROBIT models to predict bankruptcies. In addition, several studies in the past used artificial intelligence techniques to predict financial distress. In one of the earliest studies, Odom and Sharda [20] and Tam and Kiang [27] introduced BPN for predicting corporate bankruptcies. Following these studies, a number of studies further investigated the use of data mining techniques in financial distress prediction. The authors of these studies mainly tested the feasibility of BPN in bankruptcy prediction. However, BPN has many disadvantages including the need for the determination of the value of controlling parameters and the number of processing elements in the layer, as well as the danger of overfitting problem. As a result, SVM that can mitigate the limitations of BPN is recently emerging as an effective classifier for bankruptcy prediction.

Support Vector Machines (SVM)

SVM uses a linear model to implement nonlinear class boundaries by nonlinear mapping of the input vectors x into the high-dimensional feature space. A linear model constructed in the new space can represent a nonlinear boundary in the original space. In the new space, an optimal separating hyperplane is constructed [29].
Thus, SVM is known as the algorithm that finds a special kind of linear model, the *maximum margin hyperplane*. The maximum margin hyperplane gives the maximum separation between the decision classes. The training examples that are closest to the maximum margin hyperplane are called *support vectors*. All other training examples are irrelevant for defining the binary class boundaries.
SVM constructs a linear model to implement nonlinear class boundaries through the transformation of the inputs into the high-dimensional feature space. The function, $K(\mathbf{x}_i,\mathbf{x}_j)$, which is called 'kernel function', does this work. There are some different kernels for generating the inner products to construct machines with different types of nonlinear decision surfaces in the input space. Choosing among different kernels the model that minimizes the estimate, one chooses the best model. Common examples of the kernel function are the polynomial kernel $K(\mathbf{x}_i,\mathbf{x}_j)=(1+\mathbf{x}_i^T\mathbf{x}_j)^d$ and the Gaussian radial basis function (RBF) $K(\mathbf{x}_i,\mathbf{x}_j)=\exp(-1/\delta^2(\mathbf{x}_i-\mathbf{x}_j)^2)$ where d is the degree of the polynomial kernel and δ^2 is the bandwidth of the Gaussian RBF kernel [12].
As mentioned above, BPN has been widely used in the area of financial forecasting because of its broad applicability to many business problems and preeminent learning ability. On the other hand, there are no parameters to tune except the upper bound C for the non-separable cases in linear SVM [4]. Overfitting is also unlikely to occur with SVM. Overfitting may be caused by too much flexibility in the decision boundary, but the maximum hyperplane is relatively stable and gives little flexibility [30].
Although SVM has the above advantages, there are a few studies on the application of SVM in financial forecasting. Mukherjee et al. [19] showed the applicability of SVM to time-series forecasting. Tay and Cao [28] examined the predictability of financial time-series with SVMs. They showed that SVMs outperformed the BPNs on the criteria of normalized mean square error, mean absolute error, directional symmetry and weighted directional symmetry. Kim [12] applied SVM to predicting the future direction of the stock price index. In his study, SVM outperformed BPN and case-based reasoning for the prediction of the stock price index. Recently, several studies investigated the efficacy of applying SVM to bankruptcy prediction. Fan and Palaniswami [5] showed that SVM outperformed traditional classifiers for bankruptcy prediction such as DA, multi-layer perceptron, and learning vector quantization. Shin et al. [25] pointed out that the accuracy and generalization performance of SVM were better than those of BPN as the training set size got smaller. Min and Lee [17] showed that SVM outperformed LOGIT, DA, and BPN for bankruptcy prediction.

Genetic Algorithms (GA)

The genetic algorithm is a popular optimization method that attempts to incorporate ideas of natural evolution. Its procedure improves the search results by constantly trying various possible solutions with some kinds of genetic operations. In general, the process of GA proceeds as follows.
First of all, GA generates a set of solutions randomly that is called an initial population. Each solution is called a chromosome and it is usually in the form of a binary string.

After the generation of the initial population, a new population is formed that consists of the fittest chromosomes as well as offspring of these chromosomes based on the notion of survival of the fittest. The value of the fitness for each chromosome is calculated from a user-defined function. Typically, classification accuracy (performance) is used as a fitness function for classification problems.
In general, offspring are generated by applying genetic operators. Among various genetic operators, selection, crossover and mutation are the most fundamental and popular operators. The selection operator determines which chromosome will survive. In crossover, substrings from pairs of chromosomes are exchanged to form new pairs of chromosomes. In mutation, with a very small mutation rate, arbitrarily selected bits in a chromosome are inverted. These steps of evolution continue until the stopping conditions are satisfied [6,7].

Optimization of SVM using GA

Until now, researchers have studied optimization of SVM using GA in three ways. First, some studies have tried to optimize 'the kernel function and its parameters'. For example, Pai and Hong [22] used GA to optimize the free parameters used in the kernel function of SVM. The SVM model of their study used Gaussian RBF as the kernel function, and they designed the proposed system to optimize C, δ^2, ε parameters using GA. Howley and Madden [9] extended the area of optimization. Their proposed model optimized the kernel parameters (e.g. C, δ^2, d, ε) as well as the kernel function itself. Consequently, their model could present a globally optimized kernel function and its optimized parameters.
The second approach of GA-optimization of SVM is 'feature subset selection'. Feature subset selection is a method that uses only a small subset of features that prove to be relevant to the target concept. In most classification problems, the selection of an appropriate feature subset is important because it enhances classification performance by characterizing each sample more accurately, and it also reduces computational requirements. Thus, many researchers have tried to optimize the input features of SVM by using GA. For example, Lee and Byun [15] and Sun et al. [26] used this technique for image identification, and Li et al. [16] used it for cancer detection. In addition, this technique is adopted in various application areas including gear fault detection [24], abnormal key stroke detection [31], direct marketing [32], and bankruptcy prediction [18].
The final approach is 'simultaneous optimization of kernel parameters and feature subset selection'. As mentioned above, both kernel parameters and feature selection affects the classification performance of SVM. Thus, it may be more effective to optimize these factors simultaneously. Nonetheless, this is still an undiscovered area, so there are few related studies. Jack and Nandi [10] applied this technique for machinery fault detection, and Kim et al. [11] used it for network intrusion detection. Zhao et al. [33] proposed this approach to enhance protein sequence classification.
Although there have been many prior studies that optimized the various factors of SVM using GA, there is another factor to be optimized – optimal instance selection. Instance selection is the technique that selects an appropriate reduced subset of the training samples and only uses the selected subset for training. This prevents the distorted training of SVM by reducing the possibility of selecting noisy training samples as the support vectors, so it may improve classification accuracy of SVM. Due to its advantages, it has been applied to various classification techniques including neural networks [23] and case-based reasoning [2]. However, there has been no study that has introduced instance selection using GA for SVM, as far as we know. Thus, in this study, we propose a global optimization model that optimizes the selection of features, instances, and kernel parameters simultaneously by using GA.

Simultaneous Optimization of SVM using GA

This study proposes a novel SVM model whose feature selection, instance selection, and kernel parameter settings are globally optimized, in order to improve prediction accuracy of typical SVM. We employ GA to optimize these factors simultaneously. Hereafter, we call our model SOSVM - Simultaneous Optimization of SVM using GA. The detailed explanation for each step of SOSVM is presented as follows.

Phase 1. Initiation

In the first step, the system generates the initial population that would be used to find global optimum factors – feature and instance selection variables, and kernel parameters. The values of the chromosomes for the population are initiated into random values before the search process. To enable GA to find the optimal factors, we should design the structure of a chromosome as a form of binary strings. Each chromosome for SOSVM has all the information for feature selection, instance selection, and kernel parameter settings. The length of each chromosome is $m+n+12$ bits when m is the number of features and n is the number of instances. The values of the codes for feature selection and instance selection are set to '0' or '1'. '0' means the corresponding feature or instance is not selected and '1' means it is selected. The sign for feature and instance selection needs just 1 bit. As a result, $m+n$ bits are just required to implement feature and instance selection by GA. The remaining 12 bits are used for selecting appropriate kernel parameters. Similar to the study by Pai and Hong [22], we use the Gaussian radial basis function (RBF) as the kernel function of SVM. Tay and Cao [28] showed that the upper bound C and the kernel parameter δ^2 play an important role in the performance of SVM using Gaussian RBF. Setting these two parameters improperly can cause overfitting or underfitting problems. Thus, SOSVM tries to optimize these parameters using GA, and it assigns 6 bits to represent each variable. Thus, 12 bits in total are used for setting C and δ^2.

Phase 2. Training

After generating the initial population, the system performs a typical SVM process using the assigned value of the factors in the chromosomes, and calculates the performance of each chromosome. The performance of each chromosome can be calculated through the fitness function for GA. In this study, the main goal is to find the optimal or near optimal parameters that produce the most accurate prediction solution. Thus, we set the fitness function for the test data set to the prediction accuracy of the test dataset [6,13,14].

Phase 3. Genetic Operation

In the third step, a new generation of the population is produced by applying genetic operators such as selection, crossover, and mutation. According to the fitness values for each chromosome, the chromosomes whose values are high are selected and used for the basis of crossover. The mutation operator is also applied to the population with a very small mutation rate.

After the production of a new generation, phase 2 – the training process with calculation of the fitness values – is performed again. From this point, phase 2 and phase 3 are iterated again and again until the stopping conditions are satisfied. When the stopping conditions are satisfied, the genetic search finishes and the chromosome that shows the best performance in the last population is selected as the final result.

Phase 4. Checking Generalizability

Occasionally, the optimized parameters determined by GA fit quite well with the test data, but they don't fit well with the unknown data. The phenomenon occurs when the parameters fit too well with the given test data set. Thus, in the last stage, the system applies the finally selected parameters – the optimal selections of features and instances, and the optimal kernel parameters – to the hold-out (unknown) data set in order to check the generalizability of the determined factors.

The Research Design and Experiments

Application Data

The application data used in this study consists of financial ratios and the status of bankruptcy or non-bankruptcy for corresponding corporate. The data was collected from one of the largest commercial banks in Korea. The sample of bankrupt companies was 774 companies in heavy industry that filed for bankruptcy between 1999 and 2002. There were also 774 non-bankrupt companies from the same industry and period. Thus, the total size of the sample was 1548 companies.

The financial status for each company is categorized as "0" or "1" and it is used as a dependent variable. "0" means that the corporation is bankrupt, and "1" means that the corporation is solvent. For independent variables, we first generate 162 financial ratios from the financial statement from each company. Finally, we get 41 financial ratios as independent variables through the two independent sample t-test, the forward selection procedure based on logistic regression, and the opinions of the experts who are responsible for approving and managing loans in the bank. We split the data into three groups: training, test, and hold-out datasets. The portion of these groups is 60% (928 companies), 20% (310 companies) and 20% (310 companies) each.

Comparative Models

To test the effectiveness of the proposed model, we compare the result of SOSVM to the results of four different models. The first model, labeled COSVM (COnventional SVM), uses the conventional approach of SVM. This model considers all initially available features as a feature subset. That is to say, there is no special process of feature subset selection. In addition, instance selection is not considered here, so all instances are used in this model. The kernel parameters in this model are determined by varying their values to select optimal values that produce the best prediction performance.

The second model determines the optimal kernel parameters by applying GA. We call this model KPSVM (Kernel Parameter optimization for SVM by GA). Similar to COSVM, KPSVM also does not contain any function of feature selection or instance selection. Pai and Hong [22] proposed a similar model.

The third model selects relevant features using GA. This model is called FSSVM (Feature Selection for SVM by GA). Here, we try to optimize feature selection and kernel parameters by GA, but we are still unconcerned with instance selection. The studies by Jack and Nandi [10], Kim et al. [11], and Zhao et al. [33] are the examples that used this model.

The final model uses GA to select a relevant instance subset. This model is called ISSVM (Instance Selection for SVM by GA). In this model, we try to optimize instance selection and kernel parameters by GA, but we are unconcerned with feature selection.

Experimental Settings and System Development

For the controlling parameters of GA search for ISSVM and SOSVM, the population size was set at 200 organisms and the crossover and mutation rates were set at 70% and 10%. As the stopping condition, 100 generations were permitted. However, the genetic search space of KPSVM and FSSVM is much smaller than the space of ISSVM and SOSVM. Thus, we assigned 100 organisms for the population, and set the mutation rate at 15% in the case of KPSVM and FSSVM.

Table 1 – Average prediction accuracies of the models

Model	Train	Test	Hold-out	Kernel parameter	# of selected features	# of selected instances
COSVM	82.65%	-	74.52%	C=100, δ^2=25	41	928
KPSVM	84.81%	77.42%	76.45%	C=55.88, δ^2=13.04	41	928
FSSVM	82.54%	77.74%	77.42%	C=93.29, δ^2=8.76	25	928
ISSVM	84.55%	80.00%	79.35%	C=36.82, δ^2=12.65	41	492
SOSVM	81.46%	81.94%	79.68%	C=80.94, δ^2=25.85	39	701

These experiments are done by our private experimental software that is designed to perform SVM training by using parameters optimized by GA. This software is developed on a Java platform, and the class for SVM training is programmed using LIBSVM, a public software for SVM [3].

Experimental Results

In this section, the prediction performances of SOSVM and other alternative models are compared. Table 1 describes the average prediction accuracy of each model. As shown in Table 1, SOSVM achieves the higher prediction accuracy than COSVM, KPSVM, FSSVM, and ISSVM by 5.16%, 3.23%, 2.26%, and 0.33% for the hold-out data. Comparing the performance of FSSVM and ISSVM, we can find that ISSVM outperforms FSSVM by 1.93%. It may be understood that appropriate instance selection is more important than feature selection for improving prediction accuracy of SVM.
We use the two-sample test for proportions to examine whether the differences of prediction accuracy between SOSVM and other comparative algorithms are statistically significant. By applying this test, it is possible to check whether there is a difference between two probabilities when the prediction accuracy of the left-vertical methods is compared with the right horizontal methods [8]. In this test, the null hypothesis is H_0: $p_i - p_j = 0$ where i=1,...,4 and j=2,...,5, while the alternative hypothesis is H_a: $p_i - p_j > 0$ where i=1,...,4 and j=2,...,5. p_k means the classification performance of the kth method. Table 2 shows *Z values* for the pairwise comparison of the performance of the models.

Table 2 - Z values of the two sample test for proportions

	KPSVM	FSSVM	ISSVM	SOSVM
COSVM	0.560	0.846*	1.430**	1.529**
KPSVM		0.286	0.871*	0.971*
FSSVM			0.585	0.685
ISSVM				0.100

As shown in Table 2, SOSVM is better than COSVM at the 5% and better than KPSVM at the 10% statistical significance level. But, SOSVM does not outperform FSSVM and ISSVM with statistical significance.

Concluding Remarks

We have proposed a new hybrid SVM model using GA called SOSVM. Our proposed model optimizes feature selection, instance selection, and kernel parameters simultaneously. Although GA-optimization models for feature selection and kernel parameter selection of SVM have been suggested in the literature, our proposed model is designed to include 'instance selection', which reduces distorted training samples that may lead erroneous prediction. Compared to other models such as COSVM, KPSVM and FSSVM, SOSVM as well as ISSVM showed higher prediction accuracy in the empirical test for real-world bankruptcy prediction.
However, this study has some limitations. First of all, our model requires a high level of computational resources. Similar to other GA-based optimization models, SOSVM iterates the SVM training process whenever genetic evolution occurs. In particular, the search space of our model is very large, so it takes more time to get enough training. Consequently, the efforts to make SOSVM more efficient should be followed in the future. Second, the generalizability of SOSVM should be tested further. Although we apply this model to bankruptcy prediction, SOSVM can be applied to any domain that requires accurate prediction. Moreover, SOSVM did not outperform most comparative models with statistical significance in our experiment because of the insufficient sample size. Thus, it is necessary to validate the general applicability of SOSVM by applying it to other problem domains in the future.

References

[1] Altman, E.I. (1968). "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *The Journal of Finance*, Vol. 23, No. 4, pp. 589-609.

[2] Babu, T.R., and Murty, M.N. (2001). "Comparison of genetic algorithm based prototype selection schemes," *Pattern Recognition*, Vol. 34, No. 2, pp. 523-525.

[3] Chang, C.-C., and Lin, C.-J. (2001). LIBSVM: a library for support vector machines. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[4] Drucker, H., Wu, D., and Vapnik, V.N. (1999). "Support vector machines for spam categorization," *IEEE Transactions on Neural Networks*, Vol. 10, No. 5, pp. 1048-1054.

[5] Fan, A., and Palaniswami, M. (2000). "Selecting bankruptcy predictors using a support vector machine approach," *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*, pp. 354-359.

[6] Fu, Y., and Shen, R. (2004). "GA based CBR approach in Q&A system," *Expert Systems with Applications*, Vol. 26, No. 2, pp. 167-170.

[7] Han, J., and Kamber, M. (2001). *Datamining: Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann Publishers.

[8] Harnett, D.L., and Soni, A.K. (1991). *Statistical methods for business and economics*. Massachusetts, MA: Addison-Wesley.

[9] Howley, T., and Madden, M.G. (2005). "The Genetic Kernel Support Vector Machine: Description and Evaluation," *Artificial Intelligence Review*, Vol. 24, No. 3-4, pp. 379 – 395.

[10] Jack, L.B., and Nandi, A.K. (2002). "Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms," *Mechanical Systems and Signal Processing*, Vol/ 16, No. 2-3, pp. 373-390.

[11] Kim, D.S., Nguyen, H.-N., and Park, J.S. (2005). "Genetic algorithm to improve SVM based network intrusion detection system," *Proceedings of the 19th International Conference on Advanced Information Networking and Applications*, pp. 155-158

[12] Kim, K. (2003). "Financial forecasting using support vector machines," *Neurocomputing*, Vol. 55, Nos. 1-2, pp. 307-319.

[13] Kim, K. (2004). "Toward global optimization of case-based reasoning systems for financial forecasting," *Applied Intelligence*, Vol. 21, No. 3, pp. 239-249.

[14] Kim, K. (2006). "Artificial neural networks with evolutionary instance selection for financial forecasting," Expert Systems with Applications, Vol. 30, No. 3, pp. 519-526.

[15] Lee, K., and Byun, H. (2003). "A New Face Authentication System for Memory-Constrained Devices," *IEEE Transactions on Consumer Electronics*, Vol. 49, No. 4, pp. 1214-1222.

[16] Li, L., Tang, H., Wu, Z., Gong, J., Gruidl, M., Zou, J., Tockman, M., and Clark, R.A. (2004). "Data mining techniques for cancer detection using serum proteomic profiling," *Artificial Intelligence in Medicine*, Vol. 32, No. 2, pp. 71-83.

[17] Min, J.H., and Lee, Y.-C. (2005). "Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters," *Expert Systems with Applications*,Vol. 28, No. 4, pp. 603-614.

[18] Min, S.-H., Lee, J., and Han, I. (2006). "Hybrid genetic algorithms and support vector machines for bankruptcy prediction," *Expert Systems with Applications*, Vol. 31, No. 3, pp. 652-660.

[19] Mukherjee, S., Osuna, E., and Girosi, F. (1997). "Nonlinear prediction of chaotic time series using support vector machines," *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing*, pp. 511-520

[20] Odom. M., and Sharda, R. (1990). "A neural network model for bankruptcy prediction," *Proceedings of the International Joint Conference on Neural Networks*, pp. 163-168.

[21] Ohlson, J. (1980). "Financial ratios and the probabilistic prediction of bankruptcy," *Journal of Accounting Research*, Vol. 18, No. 1, pp. 109-131.

[22] Pai, P.-F., and Hong, W.-C. (2005). "Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms," *Electric Power Systems Research*, Vol. 74, No. 3, pp. 417-425.

[23] Reeves, C.R., and Taylor, S.J. (1998). "Selection of training sets for neural networks by a genetic algorithm," In Eiden, A.E., Back, T., Schoenauer M., Schwefel, H.-P.: Parallel problem-solving from nature-PPSN V. Springer. Berlin.

[24] Samanta, B. (2004). "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mechanical Systems and Signal Processing*, Vol. 18, No. 3, pp. 625-644.

[25] Shin, K.-S., Lee, T.S., and Kim, H.-j. (2005). "An application of support vector machines in bankruptcy prediction model," *Expert Systems with Applications*, Vol. 28, No. 1, pp. 127-135.

[26] Sun, Z., Bebis, G., and Miller, R. (2004). "Object detection using feature subset selection," *Pattern Recognition*, Vol. 37, No. 11, pp. 2165-2176.

[27] Tam, K.Y., and Kiang, M.Y. (1992). "Managerial applications of the neural networks: The case of bank failure predictions," *Management Science*, Vol. 38, No. 7, pp. 926-947.

[28] Tay, F.E.H., and Cao, L. (2001). "Application of support vector machines in financial time series forecasting," *OMEGA: The International Journal of Management Science*, Vol. 29, No. 4, pp. 309-317.

[29] Vapnik, V.N. (1998). *Statistical Learning Theory*. New York, NY: Wiley.

[30] Witten, I.H., and Frank, E. (2000). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. San Francisco, CA: Morgan Kaufmann Publishers.

[31] Yu, E., and Cho, S. (2004). "Keystroke dynamics identity verification: its problems and practical solutions," *Computers & Security*, Vol. 23, No. 5, pp. 428-440.

[32] Yu, E., and Cho, S. (2006). "Constructing response model using ensemble based on feature subset selection," *Expert Systems with Applications*, Vol. 30, No. 2, pp. 352-360.

[33] Zhao, X.-M., Cheung, Y.-M., and Huang, D.-S. (2005). "A novel approach to extracting features from motif content and protein composition for protein sequence classification," *Neural Networks*, Vol. 18, No. 8, pp. 1019-1028.