Support vector machines with optimal instance selection: An application to bankruptcy prediction

Hyunchul Ahn[a], Kyoung-jae Kim[b] and Ingoo Han[a]

[a] *Graduate School of Management, Korea Advanced Institute of Science and Technology*
207-43 Cheongrangri-Dong, Dongdaemun-Gu, Seoul 130-722, Korea
Tel: +82-2-958-3685, Fax: +82-2-958-3604, E-mail:hcahn@kaist.ac.kr, ighan@kgsm.kaist.ac.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: +82-2-2260-3684, E-mail:kjkim@dongguk.edu

Abstract

Building accurate corporate bankruptcy prediction models has been one of the most important research issues in finance. Recently, support vector machines (SVMs) are popularly applied to bankruptcy prediction because of its many strong points. However, in order to use SVM, a modeler should determine several factors by heuristics, which hinders from obtaining accurate prediction results by using SVM. As a result, some researchers have tried to optimize these factors, especially the feature subset and kernel parameters of SVM. But, there have been no studies that have attempted to determine appropriate instance subset of SVM, although it may improve the performance by eliminating distorted cases. Thus, in this study, we propose the simultaneous optimization of the instance selection as well as the parameters of a kernel function of SVM by using genetic algorithms (GAs). Experimental results show that our model outperforms not only conventional SVM, but also prior approaches for optimizing SVM.

Keywords:

Support vector machines; Genetic algorithms; Parameter selection; Feature selection; Instance selection

Introduction

Corporate financial distress and bankruptcy prediction is one of the major application areas of data mining. Bankruptcy prediction models have used various statistical and artificial intelligence techniques. These techniques include discriminant analysis, logistic regression, decision trees, *k*-nearest neighbor, and backpropagation (BP) neural network. Among them, the BP network 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 (Kim, 2003). 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 (Fan & Palaniswami, 2000; Min & Lee, 2005; Min et al., 2006; Shin et al., 2005).

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. Nonetheless, there have been few studies that have applied instance selection to SVM.

Instances are a collection of training examples in supervised learning and instance selection chooses a part of the data that is representative and relevant to the characteristics of all the data. Instance selection is one of popular methods for dimensionality reduction and is directly related to data reduction. Although instance selection is the most complex form of data reduction because the computationally expensive prediction methods must be invoked more often to determine the effectiveness of instance selection, we can usually remove irrelevant instances as well as noise and redundant data.

Thus, in this study, we propose a novel hybrid SVM classifier with optimization of instance subsets. This study introduces genetic algorithms (GAs) to optimize the instance selection in SVM. Our study applies the proposed model to the real-world case for bankruptcy prediction, and compares it to the typical hybrid model of GA and SVM

with optimal feature subsets and kernel parameters.
The rest of this paper is organized as follows: The next section presents the research background. Section 3 proposes the optimization model of instance selection process in SVM classifier and describes the benefits of the proposed model. Section 4 describes the application of the proposed model. Section 5 presents experimental results of the proposed and the comparative models. In the final section, conclusions and the limitations of this study are presented.

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. In addition, we review the basic concept and applications of instance selection algorithm. 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 (LR), and discriminant analysis (DA) have been employed to predict corporate bankruptcy.
Early studies by Altman (1968) used discriminant analysis to predict corporate bankruptcies. More recent research by Ohlson (1980) 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 (1990) and Tam and Kiang (1992) 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. Table 1 summarizes literatures on application of data mining techniques for bankruptcy prediction.

Table 1. Prior research on application of data mining techniques for bankruptcy prediction

Reference	ANN model	Benchmark models
Tam and Kiang, 1992	BPN	DA, LR, *k*-NN, ID3
Martin-del-Brio and Serrano-Cinca, 1993	SOM	N/A
Serrano-Cinca, 1996	SOM	N/A
Serrano-Cinca, 1997	BPN	DA, LR
Altman et al., 1994	BPN	DA
Wilson and Sharda, 1994	BPN	DA
Boritz and Kennedy, 1995	BPN	DA, Logit ,Probit
Boritz et al., 1995	BPN	DA, *k*-NN, Logit, Probit
Jo and Han, 1996	BPN	DA, *k*-NN
Lee et al., 1996	BPN	LR, DA
Jo et al., 1997	BPN	DA, *k*-NN
Kiviluoto, 1998	SOM, RBF-SOM, LVQ	DA, *k*-NN
Yang et al. 1999	PNN, BPN	DA

BPN: Backpropagation neural networks, SOM: Self-organizing map, RBF: Radial basis function, LVQ: Learning vector quantization, DA: Discriminant analysis, LR: Logistic regression, *k*-NN: *k*-nearest neighbor

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.

Instance Selection

Instance-based learning algorithms often faced the problem of deciding which instances to store for use during generalization in order to avoid excessive storage and time complexity, and to improve generalizability by avoiding noise and overfitting (Wilson & Martinez, 2000). Many researchers have addressed the problem of training data reduction and have presented algorithms for maintaining an instance base or case base in instance-based learning algorithms.
Kuncheva (1993) classified instance selection techniques (or editing techniques) into the following three categories: Condensed Nearest Neighbor rule, Generated or Modified Prototypes, and Two-Level Classifiers. The following presents some basic concepts of each category as described by prior research. A detailed explanation may be found in the references in this paper.

Condensed nearest neighbor rule: Hart (1968) made one of the first attempts to develop an instance selection rule. Hart's algorithm, the Condensed Nearest Neighbor rule, finds a subset S of the training set T such that every member of T is closer to a member of S of the same class than to a member of S of a different class. Subsequent work extended Hart's algorithm, specifically the Selective Nearest Neighbor rule (Ritter et al., 1975) and the Reduced Nearest Neighbor rule (Gates, 1972). In addition, Wilson (1972) introduced the Edited Nearest Neighbor algorithm and Tomek (1976) proposed the All k-NN method of editing.

Generated or modified prototypes: This category is composed of techniques that establish new prototypes or adjust a limited number of instances. A large group of studies within this category are implemented by ANN including feature-map classifiers, learning vector quantiziers (Kuncheva, 1995).

Two-level classifiers: This category employs two or more classifiers and allocates a part of all instances to the classifier which appears most appropriate. Tetko and Villa (1997) proposed the Efficient Partition algorithm which is used to obtain an efficient partition of noisy instances, whose distribution is proportional to the complexity of the analyzed function. This is to focus the training of ANN on the most complex and informative domains of the data set and accelerate the learning phase. They concluded that the efficiently partitioned instances enhance the predictability of ANN in comparison with a random selection of instances.

Instance selection in instance-based learning algorithms may be considered as a method of knowledge refinement and it maintains the instance-base. In this sense, some researchers proposed many instance selection algorithms for maintaining the case-base in case-based reasoning (CBR) systems. Smyth (1998) presented an approach to maintenance which is based on the deletion of harmful and redundant cases from the case-base. In addition, McSherry (2000) suggested an instance selection method in the construction of a case library in which evaluation of the coverage contributions of candidate instances are based on an algorithm called *disCover*. This algorithm reverses the direction of CBR to discover all cases that can be solved with a given case-base.

Although many different approaches have been used to address the problem of case authoring and data explosion for instance-based algorithms, there is little research on instance selection in ANN. Reeves and Taylor (1998) suggested that a GA is a promising approach to finding 'better' training data set for classification problems in radial basis function (RBF) nets. Reeves and Bush (2001) reported that the GA can also be used effectively to find a smaller subset of a 'good' training set in RBF nets for both classification and regression problems.

Support vector machine (SVM)

SVM uses a linear model to implement nonlinear class boundaries by nonlinear map-ping 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 (Vapnik, 1998).

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 (Kim, 2003).

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 (Drucker et al., 1999). 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 (Witten & Frank, 2000).

Although SVM has the above advantages, there are a few studies on the application of SVM in financial forecasting. Mukherjee et al. (1997) showed the applicability of SVM to time-series forecasting. Tay and Cao (2001) 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 (2003) 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 (2000) showed that SVM outperformed traditional classifiers for bankruptcy prediction such as DA, multi-layer perceptron, and learning vector quantization. Shin et al. (2005) 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 (2005) showed that SVM outperformed LOGIT, DA, and BPN for bankruptcy prediction.

Genetic algorithm (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 (Fu & Shen, 2004; Han & Kamber, 2001).

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 (2005) 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 (2005) 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 (2003) and Sun et al. (2004) used this technique for image identification, and Li et al. (2004) used it for cancer detection. In addition, this technique is adopted in various application areas including gear fault detection (Samanta, 2004), abnormal key stroke detection (Yu & Cho, 2004), direct marketing (Yu & Cho, 2006), and bankruptcy prediction (Min et al., 2006).

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 (2002) applied this technique for machinery fault detection, and Kim et al. (2005) used it for network intrusion detection. Zhao et al. (2005) 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. 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. 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 an optimization model of proper instances and kernel parameters for SVM classifier using GA

Optimal Instance Selection for SVM using GA

This study proposes a novel SVM model whose instance selection and kernel parameter setting are globally optimized, in order to improve prediction accuracy of typical SVM. We employ GA to optimize these factors. Hereafter, we call our model OIS – SVM with Optimal Instance Selection by GA. The detailed explanation for each step of OIS 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 – instance subset. 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 OIS has the information for instance selection and kernel parameter settings. The length of each chromosome is $n+12$ bits when n is the number of instances. The values of the codes for instance selection are set to '0' or '1'. '0' means the corresponding instance is not selected and '1' means it is selected. The sign for an instance selection needs just 1 bit. As a result, n bits are just required to implement instance selection by GA. The remaining 12 bits are used for selecting appropriate kernel parameters. Similar to the study by Pai and Hong (2005), we use the Gaussian radial basis function (RBF) as the kernel function of SVM. Tay and Cao (2001) 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, OIS 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 (Fu & Shen, 2004; Kim, 2004; Kim, 2006).

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 instances and the optimal kernel parametes – to the hold-out (unknown) data set in order to check the generalizability of the determined factors. Figure 1 presents the overall process of OIS.

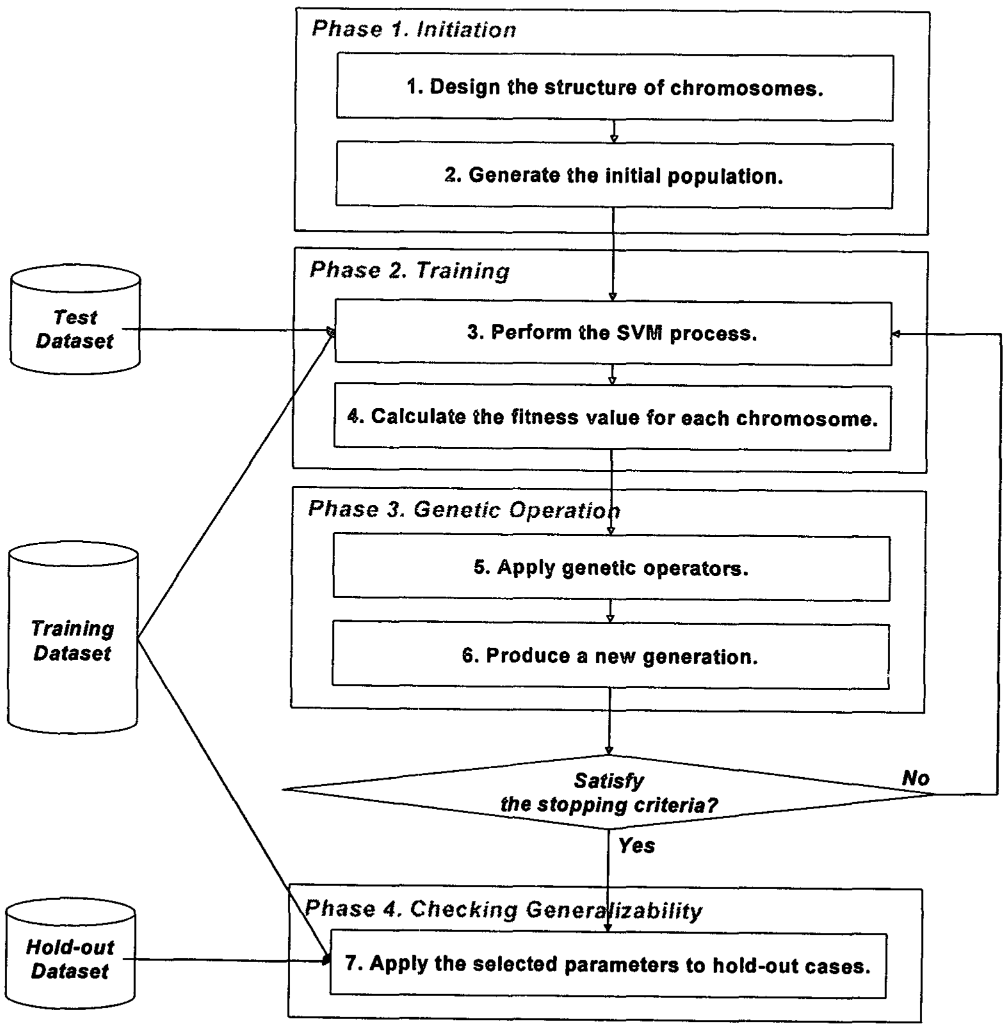

Figure 1 - The process of OIS

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 OIS to the results of three different models. The first model, labeled CON (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 OKP (SVM with Optimal Kernel Parameters by GA). Similar to CON model, OKP also does not contain any function of feature selection or instance selection. Pai and Hong (2005) proposed a similar model.

The third model selects relevant features using GA. This model is called OFS (SVM with Optimal Feature Selection 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 (2002), Kim et al. (2005), and Zhao et al. (2005) are the examples that used this model.

Research design and system development

For the controlling parameters of GA search for OIS, 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 OKP and OFS is much smaller than the space of OIS. Thus, we assigned 100 organisms for the population, and set the mutation rate at 15% in the case of OKP and OFS.

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 (Chang & Lin, 2001).

Experimental Results

In this section, the prediction performances of OIS and other alternative models are compared. Table 2 describes the average prediction accuracy of each model. As shown in Table 2, OIS achieves the higher prediction accuracy than CON OKP, and OFS by 4.83%, 2.90%, and 1.93% for the hold-out data. The difference between the performance of OIS and OFS shows that appropriate instance selection is more important than feature selection for improving prediction accuracy of SVM.

Table 2. Average prediction accuracy of the models

Model	Train	Test	Hold-out	Kernel parameter	F#[a]	I#[b]
CON	82.65%	-	74.52%	C=100, δ^2=25	41	928
OKP	84.81%	77.42%	76.45%	C=55.88, δ^2=13.04	41	928
OFS	82.54%	77.74%	77.42%	C=93.29, δ^2=8.76	25	928
OIS	84.55%	80.00%	79.35%	C=36.82, δ^2=12.65	41	492

[a] The number of selected features, [b] The number of selected instances

We use the two-sample test for proportions to examine whether the differences of prediction accuracy between OIS 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 (Harnett & Soni, 1991). In this test, the null hypothesis is H_0: $p_i - p_j = 0$ where i=1,…,3 and j=2,…,4, while the alternative hypothesis is H_a: $p_i - p_j > 0$ where i=1,…,3 and j=2,…,4. p_k means the classification performance of the kth method. Table 3 shows *Z values* for the pairwise comparison of the performance of the models.

As shown in Table 3, OIS is better than CON at the 5% and better than OKP at the 10% statistical significance level. But, OIS does not outperform OFS with statistical significance. The results show that OIS is a promising alternative model for the design of SVM classifiers.

Table 3. Z values of the two sample test for proportions

	OKP	OFS	OIS
CON	0.560	0.846*	1.430**
OKP		0.286	0.871*
OFS			0.585

* significant at the 10% level, ** significant at the 5% level

Concluding remarks

We have proposed a new hybrid SVM model using GA called OIS. Our proposed model optimizes instance selection and kernel parameters. Although GA-optimization models for feature selection and kernel parameter selection of SVM have been suggested in the previous 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 CON, OKP and OFS, OIS showed higher prediction accuracy in the empirical test for real-world bankruptcy prediction.

However, this study has some limitations. First of all, the prediction performance may be more enhanced if OIS includes feature selection process for the optimization. Although instance selection is a direct method of dimensionality reduction, feature selection effectively reduces the dimensions of feature space. Thus, simultaneous optimization of feature and instance selection may perform better than OIS and this remain a very interesting topic for further study. Second, our model requires a high level of computational resources. Similar to other GA-based optimization models, OIS 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 OIS more efficient should be followed in the future. Third, the generalizability of OIS should be tested further. Although we apply this model to bankruptcy prediction, OIS can be applied to any domain that requires accurate prediction. Moreover, it is necessary to validate the general applicability of OIS by applying it to other problem domains in the future.

References

[1] Altman, E.I.: Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. The Journal of Finance, 23(4) (1968) 589-609

[2] Altman, E.I., Macro, G., Varetto, F.: Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks. Journal of Banking and Finance, 18 (1994) 505-529

[3] Babu, T.R., Murty, M.N.: Comparison of genetic algorithm based prototype selection schemes. Pattern Recognition, 34(2) (2001) 523-525

[4] Boritz, J.E., Kennedy, D.B.: Effectiveness of neural network types for prediction of business failure. Expert Systems with Applications, 9(4) (1995) 503-512

[5] Boritz, J.E., Kennedy, D.B., de Miranda e Albuquerque, A. : Predicting corporate failure using a neural network approach. Intelligent Systems in Accounting, Finance and Management, 4 (1995) 95-111

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

[7] Drucker, H., Wu, D., Vapnik, V.N.: Support vector machines for spam categorization. IEEE Transactions on Neural Networks, 10(5) (1999) 1048-1054

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

[9] Fu, Y., Shen, R.: GA based CBR approach in Q&A system. Expert Systems with Applications, 26(2) (2004) 167-170

[10] Gates, G.W.: The reduced nearest neighbor rule. IEEE Transactions on Information Theory, 18(3) 1972 431-433

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

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

[13] Hart P.E.: The condensed nearest neighbor rule. IEEE Transactions on Information Theory, 14 (1968) 515-516

[14] Howley, T., Madden, M.G.: The Genetic Kernel Support Vector Machine: Description and Evaluation. Artificial Intelligence Review, 24(3-4) (2005) 379 - 395

[15] Jack, L.B., Nandi, A.K.: Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. Mechanical Systems and Signal Processing, 16(2-3) (2002) 373-390

[16] Jo, H., Han, I.: Integration of case-based forecasting, neural network and discriminant analysis for bankruptcy prediction. Expert Systems with Applications, 11(4) (1996) 415-422

[17] Jo, H., Han, I., Lee, H.: Bankruptcy prediction using case-based reasoning, neural network and discriminant analysis. Expert Systems with Applications, 13(2) (1997) 97-108

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

[19] Kim, K.: Financial forecasting using support vector machines. Neurocomputing, 55(1-2) (2003) 307-319

[20] Kim, K.: Toward global optimization of case-based reasoning systems for financial forecasting. Applied Intelligence, 21(3) (2004) 239-249

[21] Kim, K.: Artificial neural networks with evolutionary instance selection for financial forecasting. Expert Systems with Applications, *30(3) (2006) 519-526*

[22] Kiviluoto, K.: Predicting bankruptcies with the self-organizing map. Neurocomputing, 21(1-3) (1998) 203-224

[23] Kuncheva, L.I.: 'Change-glasses' approach in pattern recognition. Pattern Recognition Letters, 14 (1993) 619-623

[24] Kuncheva, L.I.: Editing for the k-nearest neighbors rule by a genetic algorithm. Pattern Recognition Letters, 16(8) (1995) 809-814

[25] Lee, K., Byun, H.: A New Face Authentication System for Memory-Constrained Devices. IEEE Transactions on Consumer Electronics, 49(4) (2003) 1214-1222

[26] Lee, K.C., Han, I., Kwon, Y.: Hybrid neural network models for bankruptcy predictions. Decision Support Systems, 18 (1996) 63-72

[27] Li, L., Tang, H., Wu, Z., Gong, J., Gruidl, M., Zou, J., Tockman, M., Clark, R.A.: Data mining techniques for cancer detection using serum proteomic profiling. Artificial Intelligence in Medicine, 32(2) (2004) 71-83

[28] Martin-del-Brio, B., Serrano-Cinca, C.: Self-organizing neural networks for the analysis and representation of data: Some financial cases. Neural Computing & Applications, 1 (1993) 193-206

[29] McSherry D.: Automating case selection in the construction of a case library. Knowledge Based Systems, 13(2-3) (2000) 133-140

[30] Min, J.H., Lee, Y.-C.: Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert Systems with Applications, 28(4) (2005) 603-614

[31] Min, S.-H., Lee, J., Han, I.: Hybrid genetic algorithms and support vector machines for bankruptcy prediction. Expert Systems with Applications, (2006) Forthcoming

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

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

[34] Ohlson, J.: Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, 18(1) (1980) 109-131

[35] Pai, P.-F., Hong, W.-C.: Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. Electric Power Systems Research, 74(3) (2005) 417-425

[36] Reeves, C.R., Bush, D.R.: Using genetic algorithms for training data selection in RBF networks. In Liu H., Motoda H.: Instance selection and construction for data mining. Kluwer Academic Publishers. Massachusetts (2001)

[37] Reeves, C.R., Taylor, S.J.: 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 (1998)

[38] Ritter, G.L., Woodruff, H.B., Lowry, S.R., Isenhour, T.L.: An algorithm for a selective nearest neighbor decision rule. IEEE Transactions on Information Theory, 21(6) (1975) 665-669

[39] Samanta, B.: Gear fault detection using artificial neural networks and support vector machines with genetic algorithms. Mechanical Systems and Signal Processing, 18(3) (2004) 625-644

[40] Serrano-Cinca, C.: Self organizing neural networks for financial diagnosis. Decision Support Systems, 17 (1996) 227-238

[41] Serrano-Cinca, C.: Feedforward neural networks in the classification of financial information. The European Journal of Finance, 3(3) (1997) 183-202

[42] Shin, K.-S., Lee, T.S., Kim, H.-j.: An application of support vector machines in bankruptcy prediction model. Expert Systems with Applications, 28(1) (2005) 127-135

[43] Smyth, B.: Case-base maintenance. Proceedings of the 11th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems (1998) 507-516

[44] Sun, Z., Bebis, G., Miller, R.: Object detection using feature subset selection. Pattern Recognition, 37(11) (2004) 2165-2176

[45] Tam, K.Y., Kiang, M.Y.: Managerial applications of the neural networks: The case of bank failure predictions. Management Science, 38(7) (1992) 926-947

[46] Tay, F.E.H., Cao, L.: Application of support vector machines in financial time series forecasting. OMEGA: The International Journal of Management Science, 29(4) (2001) 309-317

[47] Tetko, I.V., Villa, A.E.P.: Efficient partition of learning data sets for neural network training. Neural Networks, 10(8) (1997) 1361-1374

[48] Tomek, I.: An experiment with the edited nearest neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics, 6(6) (1976) 448-452

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

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

[51] Wilson, R.L., Sharda. R.: Bankruptcy prediction using neural networks. Decision Support Systems, 11 (1994) 545-557

[52] Wilson, D.L.: Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics, 2(3) (1972) 408-421

[53] Wilson, D.R., Martinez, T.R.: Reduction techniques for instance-based learning algorithms. Machine Learning, 38 (2000) 257-286.

[54] Yang, Z.R., Platt, M.B., Platt, H.D.: Probabilistic neural networks in bankruptcy prediction. Journal of Business Research, 44 (1999) 67-74
[55] Yu, E., Cho, S.: Keystroke dynamics identity verification: its problems and practical solutions. Computers & Security, 23(5) (2004) 428-440
[56] Yu, E., Cho, S.: Constructing response model using ensemble based on feature subset selection. Expert Systems with Applications, 30(2) (2006) 352-360
[57] Zhao, X.-M., Cheung, Y.-M., Huang, D.-S.: A novel approach to extracting features from motif content and protein composition for protein sequence classification. Neural Networks, 18(8) (2005) 1019-1028