Combining genetic algorithms and support vector machines for bankruptcy prediction

Sung-Hwan Min[a*], **Jumin Lee**[a] **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*

Abstract

Bankruptcy prediction is an important and widely studied topic since it can have significant impact on bank lending decisions and profitability. Recently, support vector machine (SVM) has been applied to the problem of bankruptcy prediction. The SVM-based method has been compared with other methods such as neural network, logistic regression and has shown good results. Genetic algorithm (GA) has been increasingly applied in conjunction with other AI techniques such as neural network, CBR. However, few studies have dealt with integration of GA and SVM, though there is a great potential for useful applications in this area. This study proposes the methods for improving SVM performance in two aspects: feature subset selection and parameter optimization. GA is used to optimize both feature subset and parameters of SVM simultaneously for bankruptcy prediction.

Keywords: Support vector machines; Bankruptcy prediction; Genetic algorithms

1. Introduction

Bankruptcy prediction is an important and widely studied topic since it can have significant impact on bank lending decisions and profitability. While companies have tried to require funding to settle their business stably in this situation, banks and other investment companies become conservative to investment for the non-first class company. However the best way for each other is not to be conservative but correctly to select healthy companies and invest them. Therefore the correctness of credit evaluation and timely investment should be emphasized more than before.

Bankruptcy is not an abrupt occurrence. It is from accumulation of management fault causes (lack management talent, economy's structural causes, lack of funding etc.) These are information and symptoms of bankruptcy prediction. Many economic researchers has developed these signs such as Interest Cover, CFAR(Cash Flow Adequacy Ratio) which the global credit evaluation organization, Fitch IBCA mentions.

Statistical methods and data mining techniques have been used for the more accurate prediction models. The former are such as regression, discriminant analysis, logistic models, factor analysis etc. The latter are such as decision trees, neural networks, fuzzy logic, genetic algorithm, SVM etc.

Bankruptcy prediction has been an important issue in the accounting and finance and challenged by the prediction models. The prediction is a kind of binary decision, in terms of two-class pattern recognition problem. Beaver (1966) originally proposed the univariate analysis on financial ratios to predict the problem and many researches have followed to improve the decision with a variety of statistical methodologies. Linear discriminant analysis[1, 2], multiple regression [23], logistic regression [11, 25, 26] have been typically used for this purpose. However strict assumptions of the traditional statistics such as the linearity, normality, independence among predictor variables and pre-existing functional form relating the criterion variable and predictor variable made limitation at application in the real world.

Recent AI approaches are inductive learning [14, 30], Case-based Reasoning (CBR) [6, 7], and ANNs [5, 9, 18, 38]. In AI approach, ANNs are powerful tools for pattern recognition and pattern classification due to their nonlinear non-parametric adaptive-learning properties. ANNs have been used successfully for many financial problems. Moreover hybrid NN models for bankruptcy with statistical and inductive learning methods [22], SOFM [21] have shown great results.

Recently SVM which is developed by Vapnik [37] is one of the methods that is receiving increasing attention with remarkable results. The main difference between ANN and SVM is the principle of risk minimization. While ANN implement empirical risk minimization to minimize the error on the training data, SVM implemented the principle of Structural Risk

[*] Corresponding author: Tel. 822-958-3131 Fax. 822-958-3604.
E-mail address: shmin @kgsm.kaist.ac.kr(Sung-Hwan Min), leejumin@kgsm.kaist.ac.kr(Jumin Lee)

Minimization by constructing an optimal separating hyper plane in the hidden feature space, using quadratic programming to find a unique solution. The difference leads better performance for SVMs than ANNs. Originally SVMs were developed for pattern recognition problems and have been used for isolated handwritten digit recognition [28], text categorization [19], speaker identification [27] and mechanical system [17]. SVMs has yielded excellent generalization performance or significantly better than that of competing methods on the problems. In financial applications time series prediction such as stock price indexing [8, 20, 34, 35], and classification such as credit rating [15], bankruptcy [12, 36] are main areas with SVMs.

On the other side, hybrid models also have advanced with these single prediction models. One of the popular hybrid models is using Genetic algorithm (GA). GA has been increasingly applied in conjunction with other AI techniques such as ANN, case based reasoning. However, few studies have dealt with integration of GA and SVM, though there is a great potential for useful applications in this area. This paper focuses on the improvement of SVM-based method by means of the integration of GA and SVM.

This study presents the methods for improving SVM performance in two aspects: feature subset selection and parameter optimization. GA is used to optimize both feature subset and parameters of SVM simultaneously for bankruptcy prediction. This paper applies the proposed GA- SVM model to bankruptcy prediction problem using real data set from Korea companies.

The rest of this paper is organized as follows: The next section describes background. Section 3 presents the proposed model. Section 4 explains the results of the evaluation experiment. The final section presents the summary and future research issue.

2. Research Background

2.1 SVM

Support Vector Machine (SVM) developed by Vapnik [37] implemented the principle of Structural Risk Minimization by constructing an optimal separating hyper plane $\mathbf{w} \bullet \mathbf{x} + b = 0$.

To the optimal hyper plane: $\{\mathbf{x} \in S : (\mathbf{w}, \mathbf{x}) + b = 0\}$, the norm of the vector w need to be minimized, in the other hand, the margin $1/\|\mathbf{w}\|$ should be maximized between two classes.

$$\min_{i=1,\dots n} |(\mathbf{w}, \mathbf{x}) + b| = 1\}$$

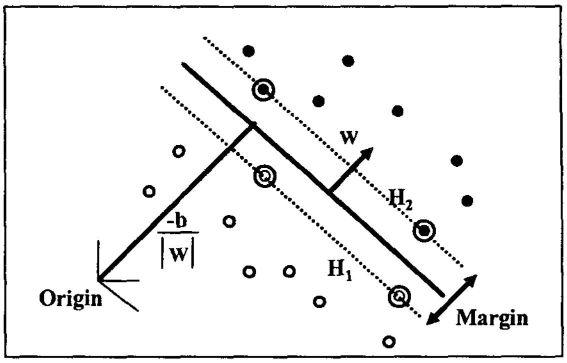


Fig.1. Linear separating hyperplanes for the separable case (The support vectors are circled)

The solution for a typical two case in linear cases has the form as shown in Figure1. Those circled points are called "support vectors" for which $y_i(x_i \cdot \mathbf{w}) + b = 1$ holds and which are confining the margin the moving of any of them will change the hyper plane normal vector w.

In non linear case, we first mapped the data to some other Euclidean space H, using a mapping, $\Phi : \mathbf{R}^d \mapsto \mathrm{H}$. Then instead of the form of dot products, "kernel function" K such that $K(\mathbf{x}_i, \mathbf{y}_i) = \Phi(\mathbf{x}_i) \bullet \Phi(\mathbf{x}_j)$. There are several Kernel functions.

Simple dot product: $K(\mathbf{x}, \mathbf{y}) = \mathbf{x} \bullet \mathbf{y}$

Vovk's polynomial: $K(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \bullet \mathbf{y} + 1)^p$

Radial Basis Function (*RBF*): $K(\mathbf{x}, \mathbf{y}) = e^{-\|\mathbf{x}-\mathbf{y}\|}$

Two Layer Neural Network: $K(\mathbf{x}, \mathbf{y}) = \tanh(k\mathbf{x} \bullet \mathbf{y} - \delta)$

Using a dual problem, the quadratic programming problems can be re-written as

$$Q(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

$$\textit{subject to} \quad 0 \le \alpha_i \le C \quad \sum_{i=1}^{l} \alpha_i y_i = 0$$

with decision function $f(\mathbf{x}) = \mathrm{sgn}(\sum_{i=1}^{l} \mathrm{y}_i \alpha_i k(x, x_i) + b))$ ·

In this paper, we define the bankruptcy problem as a non linear problem and use RBF kernel to optimize the hyper plan.

2.2 Genetic Algorithm

Genetic Algorithm (GA) is an artificial intelligence procedure based on the theory of natural selection and evolution. GA uses the idea of survival of the fittest by progressively accepting better solutions to the problems. It is inspired by and named after biological processes of inheritance, mutation, natural selection, and the genetic crossover that occurs when parents mate to produce offspring [13]. GA differs from conventional non-linear optimization techniques in that it search by maintaining a population (or data base) of solutions from which better solutions are created rather than making incremental changes to a single solution to the problem. GA simultaneously possesses a large amount of candidate solutions to a problem, called population. The key feature of a GA is the manipulation of a population whose individuals are characterized by possessing a chromosome.

Two important issues in GA are the genetic coding used to define the problem and the evaluation function, called the fitness function. Each individual solution in GA is represented by a string called the chromosome. Initial solution population could be generated randomly, which evolve to the next generation by genetic operators such as selection, crossover and mutation. The solutions coded by strings are evaluated by the fitness function. Selection operator allows strings with higher fitness to appear with higher probability in the next generation [16, 24]. Crossover is performed between two selected individuals, called parents, by exchanging parts of their strings, starting from a randomly chosen crossover point. This operator tends to enable to the evolutionary process to move toward promising regions of the search space. Mutation is used to search further space of problem and to avoid local convergence of the GA [33].

GA has been extensively researched and applied to many combinatorial optimization problems. Furthermore GA has been increasingly applied in conjunction with other AI techniques such as neural network, CBR. Various problems of neural network design have been optimized using GA. GA has been also used in conjunction with CBR to select relevant input variables and tune the parameters of CBR [4]. Few studies have dealt with integration of GA and SVM, though there is a great potential for useful applications in this area.

3. Hybrid GA-SVM Model

This study presents the methods for improving SVM performance in two aspects: feature subset selection and parameter optimization. GA is used to optimize both feature subset and parameters of SVM simultaneously for bankruptcy prediction.

3.1. Optimizing Feature Subset

Feature subset selection is essentially an optimization problem, which involves searching the space of possible features to find one that is optimum or near-optimal with respect to a certain performance measures such as accuracy. In classification problem, the selection of features is important for many reasons: good generalization performance, running time requirements and constraints imposed by the problem itself.

In the literature there are known two general approaches to solve the feature selection problem: The filter approach and the wrapper approach [32]. The distinction made depending on weather feature subset selection is done independently of the learning algorithm used to construct the classifier (i.e., filter) or not (i.e., wrapper). In the filter approach feature selection is performed before applying the classifier to the selected feature subset. The filter approach is computationally more efficient than a wrapper approach. Wrapper approach train the classifier system with a given feature subset as an input and estimate the classification error using a validation set. Although this is a slower procedure, the features selected are usually more optimal for the classifier employed.

In bankruptcy prediction problem, feature subset selection plays an important role on the performance of prediction. Furthermore its importance increases when the number of features is large. This paper seeks to improve SVM based bankruptcy prediction model. We propose the GA as the method of feature subset selection in the SVM system.

This paper uses the wrapper approach to select optimal feature subset of the SVM model using GA.

3.2. Optimizing the parameters of SVM

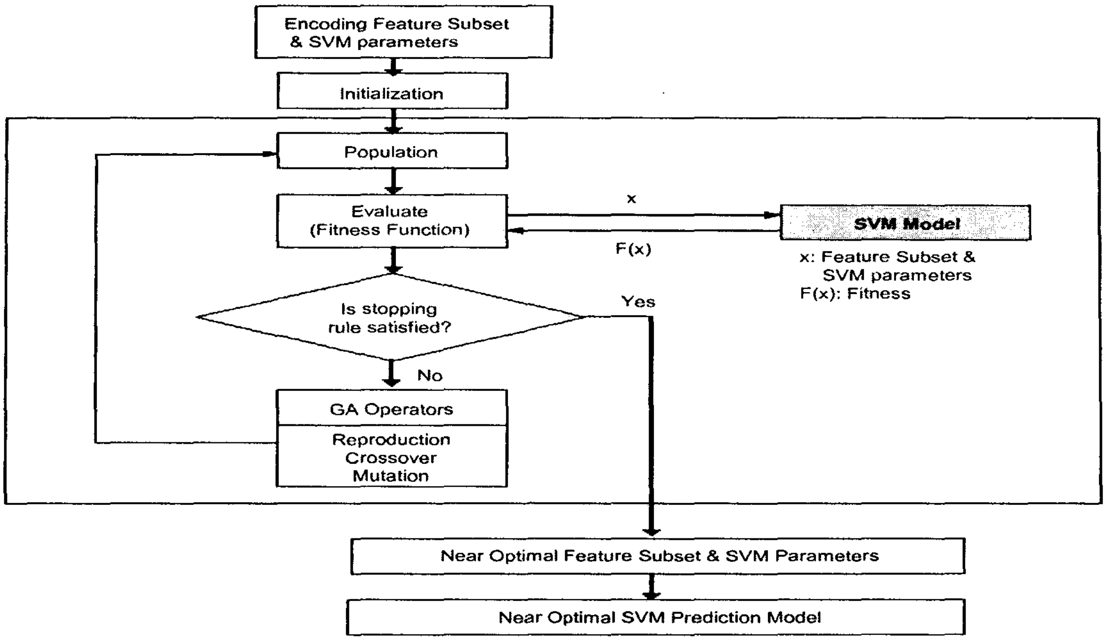
One of the big problems in SVM is the selection of the value of parameters that will allow good performance. Selecting an appropriate value for parameters of SVM plays an important role on the performance of SVM. But, it is not known beforehand which values are the best for one problem. Optimizing the parameters of SVM is crucial for the best prediction performance.

This paper proposes the GA as the method of optimizing parameters of SVM. In this paper, the radial basis function (RBF) is used as the kernel function for bankruptcy prediction. There are two parameters while using RBF kernels: C andδ^2. These two parameters play an important role in the performance of SVMs [34]. In this study, C and $\delta^2$are encoded as binary strings and optimized by GA.

3.3. Simultaneous optimization of SVM using GA

In general, the choice of the feature subset has an influence on the appropriate kernel parameters and vice versa. Therefore feature subset and parameters of SVM need to be optimized simultaneously for the best prediction performance.

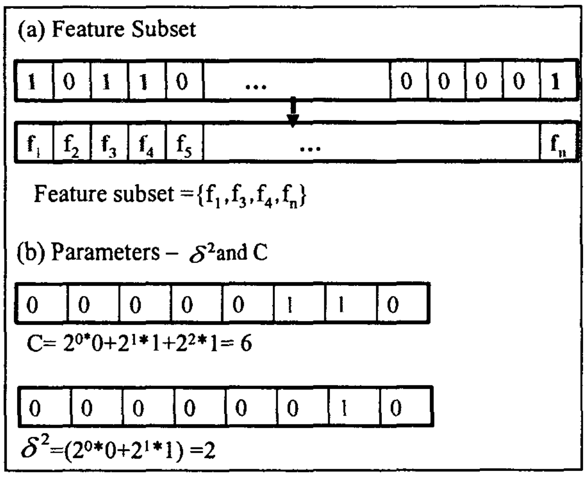
Figure 2 shows the overall procedure of the proposed model which optimizes both feature subset and parameters of SVM simultaneously for bankruptcy prediction. The procedure starts with the randomly selected chromosomes which represent feature subset and parameters of SVM. Each new chromosome is evaluated by sending it to the SVM model. The SVM model uses the feature subset and parameters in order to obtain the performance measure (e.g. hit ratio). This performance measure is used as the fitness function and is evolved by GA.



Encoding Feature Subset & SVM parameters
Initialization
Population
Evaluate (Fitness Function)
x
F(x)
SVM Model
x: Feature Subset & SVM parameters
F(x): Fitness
Is stopping rule satisfied?
Yes
No
GA Operators
Reproduction
Crossover
Mutation
Near Optimal Feature Subset & SVM Parameters
Near Optimal SVM Prediction Model

Fig.2. Overall Procedure of GA-SVM

The chromosomes for feature subset are encoded as binary strings standing for some subset of the original feature set list. Each bit of the chromosome represents whether the corresponding feature is selected or not. 1 in each bit means the corresponding feature is selected, whereas 0 means it is not selected. The chromosomes for parameters of SVM are encoded as 16-bit string which consists of 8-bit standing for C and 8-bit standing for δ^2. Figure 3 shows examples of encoding for GA.



(a) Feature Subset
1 0 1 1 0 ... 0 0 0 0 1
f_1 f_2 f_3 f_4 f_5 ... f_n
Feature subset = $\{f_1, f_3, f_4, f_n\}$
(b) Parameters – δ^2 and C
0 0 0 0 0 1 1 0
C= $2^0*0+2^1*1+2^2*1= 6$
0 0 0 0 0 0 1 0
$\delta^2=(2^0*0+2^1*1) =2$

Fig. 3. Examples of Encoding for GA

Each of the selected feature subsets and parameters is evaluated using SVM. This process is iterated until the best feature subset is and values of parameters found.

The data set is divided into a training set and a validation portion. Training set (T) consists of both T_1 and T_2.

GA evolves a number of populations. Each population consists of sets of features of a given size and the values of parameters. The fitness of an individual of the population is based on the performance of SVM. SVM is trained on T_1 using only the features of the individual and the values of parameters of the individual. The fitness is the SVM error over T_2. At each generation new individuals are created and inserted into the population by selecting fit parents which are mutated and recombined.

The fitness function is represented mathematically as follows:

$$Fitness = \frac{\sum_{i=1}^{n} H_i}{n}$$

where H_i is 1 if actual output equal to the predicted value of the SVM model, otherwise H_i is zero.

During the evolution, the simple crossover operator (traditional 1-point crossover) is used. Mutation operator just flips a specific bit. With elite survival strategy, we reserve elite not only between generations but also in the operation of crossover and mutation so that we can obtain all the benefit of GA operation. The details of the proposed model in an algorithmic form are explained in Table 1.

Table1. Step of GA-SVM

```
Step 1. Define the string (or chromosome)
        $V_{1i}$ = (s,t,...,r)(Features of SVM are encoded into chromosomes)
        $V_{2i}$  (Parameters of SVM are encoded into chromosomes)
Step 2. Define population size ($N_{pop}$), probability of crossover(Pc) and probability of mutation(Pm).
Step 3. Generate binary coded initial population of $N_{pop}$ chromosomes randomly.
Step 4. While stopping condition is false, do Step 4- 8.
Step 5. Decode $j_{th}$ chromosome ($j = 1,2, \ldots, N_{pop}$) to obtain the corresponding feature subset $V_{1i}$ and
            parameters $V_{2i}$
Step 6. Apply $V_{1j}$ and $V_{2j}$ to the SVM model to compute the output, $O_k$.
Step 7. Evaluate fitness, $F_j$ of the $j_{th}$ chromosome using $O_k$
            (Fitness function: Average predictive accuracy)
Step 8. Calculate total fitness function of population
        $TF = \sum_{i=1}^{N_{pop}} F_i(V^1{}_i, V^2{}_i)$
Step 9. Reproduction
  9.1 Compute $q_i = F_i\ (V_i)/TF$
  9.2 Calculate cumulative probability
  9.3 Generate random number r between [0, 1]. If r<q1 ,
          then select first string ($v_1$), otherwise, select $j_{th}$ string such that qi-1 <r<$q_j$
Step 10. Generate offspring population by performing crossover and mutation on parent pairs
  10.1 Crossover: Generate random number r between [0, 1] for a new string.
           If r< Pc , then operate crossover
      10.2 Mutation: Generate random number r between [0, 1] and select the bit for mutation randomly. If r1<
           Pm , then operate mutation for the bit.
Step 11. Stop the iterative step when the terminal condition is reached.
```

4. Research data and experiments

The research data used in this study is obtained from the commercial bank in Korea. The data set contains externally non-audited 614 medium-size light industry firms.

Among cases, 307 companies are bankrupt which filed for bankruptcy from 1999 to 2002.

Initially 32 financial ratios categorized as stability, profitability, growth, activity and cash flow are investigated through literature review and basic statistical method.

Using 4 feature sets with various parameter set, we experiment pure SVM to observe variety characteristics. Out of total 32 financial ratios 4 feature subsets are selected for experiment. The selected variables and feature subsets are shown Table 2. In Table 2, 32FS represents all financial ratios. 30FS means 30 financial ratios which is selected by independent-samples t-test between each financial ratio as an input variable and bankrupt or non-bankrupt as an output variable. 12FS and 6FS represent feature subset selected by logistic regression stepwise and MDA stepwise method respectively.

Table 2. Variables and Feature Subset

Category	Features	Feature Subset for model comparison				Selected by GA-SVM
		6FS	12FS	30FS	32FS	
Stability	Quick Ratio			O	O	
	Debt/Total Asset		O	O	O	
	Debt Repayment Coefficient			O	O	
	Debt Ratio			O	O	O
	Equity Capital Ratio		O	O	O	
	Debt/Total Asset		O	O	O	O
	Cash Ratio	O		O	O	
Profitability	Financial Expenses to Sales			O	O	
	Operating Income/Net Interest Expenses			O	O	
	Financial Expenses to Debt ratio	O	O	O	O	O
	Net Financing Cost/Sales			O	O	
	Time interest earned(Interest Cover)		O	O	O	O
	Ordinary Income of Total Asset		O	O	O	O
	Return on Total Asset		O	O	O	
	(Operation Profit + non Operation Profit)/Capital	O	O	O	O	O
	Net Income/Capital			O	O	
	EBIT/Interest Cost		O	O	O	
	EBITDA/Interest Cost			O	O	O
Growth	Sales Increase Ratio	O			O	O
	Growth Rate of Sales	O	O	O	O	O
	Net Profit Increase Rate				O	
Activity	Inventory Change/Sales			O	O	O
	Account Receivable Change/Sales	O	O	O	O	
	Working capital change/Sales			O	O	O
	Operating asset Change/Sales			O	O	
Cash Flow	Cash Operational Income/Debt			O	O	
	Cash Operational Income/Interest Expenses			O	O	O
	Debt Service Coverage Ratio		O	O	O	
	Cash flow from operating activity/Debt			O	O	
	Cash flow from operating activity/ Interest Expenses			O	O	
	Cash flow after interest payment/Debt			O	O	
	Cash flow after interest payment/Interest Expenses			O	O	O

We use the term, "GA-SVM" model as the proposed model which is simultaneous optimization of SVM using GA. The data set for GA-SVM is separated by 2 parts: training set, and a validation set. The ratios are about 0.7, 0.3. The training data for neural network and GA-SVM is divided into two portions again: one is for training model and the other is for avoiding over fitting. Additionally, to evaluate the effectiveness of the proposed model, we compare four different models with arbitrarily selected values of parameters and given feature subset. The first model, labeled LR, uses logistic regression. The second model, labeled NN, uses neural network and the third model, labeled Pure SVM means SVM.

5. Experiment results

5.1 Sensitivity of Pure SVM to feature sets and parameters

Table 3 shows the classification accuracies of various parameters in SVM using various feature subsets. The experimental results shows that the prediction performance of SVM is sensitive to not only various feature subset but also various parameters.

Table 3. Classification accuracies (%) of various parameters in pure SVM using various feature subset

	δ^2													
	1		10		30		50		80		100		200	
	Tr	Val	Tr	Val	Tr	Val	Tr	Val	Tr	Val	Tr	Val	Tr	Val
C	6FS													
1	80.05	70.71	75.48	71.21	75.24	69.19	75.48	68.18	75.00	69.19	72.12	70.20	62.50	57.58
10	87.74	70.20	77.40	70.71	76.44	70.71	81.25	74.75	75.24	72.22	75.72	71.72	74.76	69.19
30	92.31	64.65	79.09	70.20	75.96	69.70	76.44	70.20	75.72	70.71	75.96	71.72	75.24	71.72
50	94.71	63.64	79.81	69.70	75.72	68.69	76.20	69.19	76.68	70.20	76.20	70.71	75.96	72.22
70	95.67	64.65	80.53	70.71	75.48	69.19	76.20	69.70	76.68	70.20	76.68	70.20	75.72	71.21
90	95.91	64.14	81.49	70.71	75.72	69.19	76.20	68.69	76.20	69.70	76.44	70.71	76.20	70.71
100	95.67	64.65	81.49	70.20	76.20	69.19	76.20	68.69	75.96	69.70	76.44	70.71	75.96	70.20
150	96.39	66.16	82.69	71.72	77.16	69.19	75.00	69.19	75.96	69.70	75.96	69.70	76.92	70.20
200	97.36	64.14	82.93	70.71	78.13	69.70	75.72	68.69	76.20	69.19	75.96	69.70	76.44	71.21
250	98.08	63.13	83.65	71.21	79.09	69.70	75.48	69.19	75.96	69.19	76.20	69.19	75.96	70.20
	12FS													
1	79.09	69.19	75.96	68.69	75.24	66.16	73.80	66.16	73.80	64.14	73.32	64.65	68.99	61.62
10	81.73	72.73	78.13	67.17	76.68	67.17	76.20	67.68	76.44	68.69	75.72	68.69	75.72	67.17
30	83.89	70.20	79.09	68.18	77.88	67.68	77.88	67.68	76.92	67.17	76.20	67.68	75.72	67.17
50	86.06	69.70	78.61	69.19	78.61	67.17	77.88	67.68	78.13	68.18	77.40	67.17	76.44	67.68
70	88.22	70.71	79.57	68.69	78.13	67.17	78.13	67.17	78.37	66.67	78.37	66.67	76.92	67.17
90	88.94	70.20	79.81	70.20	77.88	67.17	78.13	67.68	78.61	66.67	78.13	67.17	77.64	68.18
100	89.66	70.71	79.57	69.70	77.88	67.68	78.61	67.68	78.37	67.17	78.37	67.17	77.64	67.17
150	90.14	69.70	79.81	69.19	77.64	68.69	78.37	67.68	78.37	68.18	78.61	68.18	78.37	67.17
200	90.38	66.67	79.57	69.70	77.40	69.19	77.88	68.69	78.37	67.68	78.37	68.18	78.85	67.17
250	91.59	66.16	80.29	68.69	77.40	69.70	77.40	68.69	79.09	68.18	78.61	67.68	78.85	67.68
	30FS													
1	81.73	70.20	76.20	70.20	74.52	69.19	73.32	66.16	72.36	66.67	72.12	65.66	71.39	64.14
10	91.35	69.19	78.61	68.69	75.72	70.71	75.96	71.21	75.72	70.20	75.24	69.70	74.76	70.20
30	95.43	64.14	81.97	68.69	76.44	67.68	76.20	70.71	75.72	69.70	75.96	70.20	75.72	69.70
50	97.60	64.14	81.97	68.18	78.85	68.18	76.20	67.68	76.44	70.20	76.20	70.20	75.72	70.20
70	98.32	64.65	82.69	69.19	79.09	69.70	76.92	67.68	76.44	69.70	76.44	69.70	75.48	69.19
90	98.56	65.15	83.17	70.20	79.57	68.69	77.40	67.68	76.20	68.18	76.68	69.70	75.72	68.69
100	98.80	64.65	83.17	70.20	79.33	68.69	77.64	68.18	77.16	68.18	76.44	69.70	75.72	68.69
150	99.28	65.66	84.86	71.72	81.49	68.18	78.85	69.19	77.16	68.18	76.92	68.18	75.96	68.69
200	99.52	67.17	85.58	70.71	81.73	68.69	79.09	69.19	76.92	67.68	76.92	68.69	76.44	69.19
250	99.76	64.65	86.30	69.70	82.45	68.69	79.33	68.69	77.88	67.68	77.16	68.18	76.68	68.69
	32FS													
1	82.45	72.22	76.68	70.71	74.76	69.70	73.32	66.16	72.36	66.67	71.88	65.66	71.63	64.14
10	93.27	66.67	78.85	68.69	76.20	69.70	75.96	70.71	75.72	70.71	74.52	68.69	74.76	69.70
30	96.88	66.16	81.25	69.19	76.92	69.19	76.20	69.19	76.20	69.19	76.20	71.21	75.48	69.70
50	98.56	64.14	83.17	70.20	78.61	69.70	75.72	68.18	76.44	69.19	76.20	68.69	75.72	70.71
70	99.04	63.64	84.13	69.70	79.57	69.70	76.68	69.19	76.20	68.69	76.20	69.19	76.44	70.71
90	99.28	63.13	85.10	70.20	80.05	68.69	77.16	69.70	75.48	68.69	75.96	68.18	75.96	70.20
100	99.28	63.64	85.34	70.20	80.53	68.18	77.40	69.70	75.96	68.69	75.72	67.68	76.20	69.19
150	99.76	62.63	86.54	71.21	81.25	69.19	79.57	69.70	76.44	68.69	76.44	68.69	76.20	68.69
200	99.76	64.65	87.02	71.21	81.49	69.19	80.53	68.18	77.40	67.68	76.68	68.18	76.20	68.18
250	99.76	64.14	88.22	70.20	82.69	68.69	80.53	68.69	79.09	68.69	77.16	67.68	76.20	68.18

In Fig.4, the best prediction with δ^2=10 and C=50 using 6FS on the validation set is poor using 12FS. In that, this results shows that simultaneous optimization of feature set and parameters is needed for the best prediction.

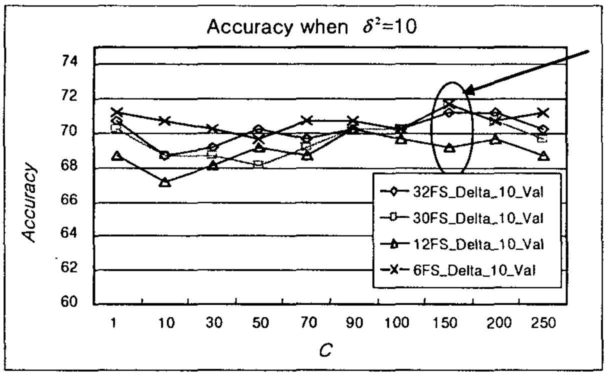


Fig. 4 Accuracy of Validation Set when δ^2 =10

Figure 5 shows one of the results of SVM with 30FS where δ^2 is fixed at 30 as C increases. As Tay and Cao [34] mentioned, we can observe that a large value for C would over-fit the training data.

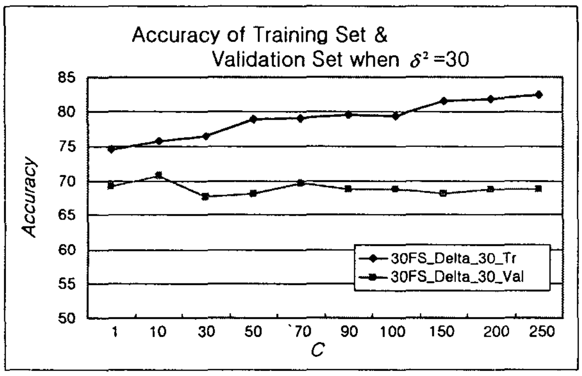


Fig. 5 Accuracy of 30FS's Training Set & Validation Set when δ^2=30

Fig.6 shows the result of SVM with 12FS where C is fixed on 70 and δ^2 increases. We can observe that a small value for δ^2 would over-fit the training data and δ^2 plays an important role on generalization performance of SVM. These result also support Tay and Cao[34]

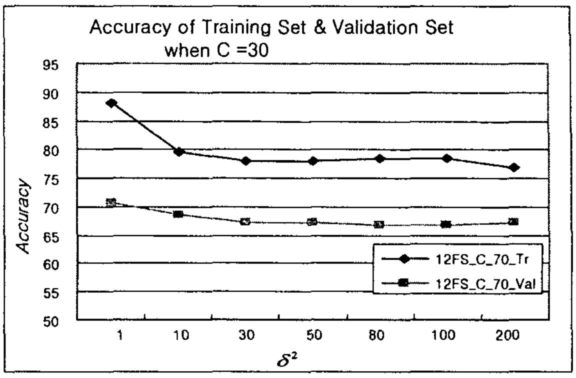


Fig. 6 Accuracy of 30FS's Training Set & Validation Set when C=3

5.2 Results of GA-SVM

Table 3 shows feature subset selected by GA. Table 4 describes the average prediction accuracy of each model. In Pure SVM, we use the best result on the validation set out of results of Table 3. In Table 4, the proposed model shows better performance than the other models. The McNemar tests are used to examine whether the proposed model significantly outperforms the other models. This test is a nonparametric test for two related samples using the chi-square distribution. The McNemar test assesses the significance of the difference between two dependent samples when the variable of interest is a dichotomy.

It is useful for detecting changes in responses due to experimental intervention in "before-and-after" designs [31].

Table 4. Average prediction accuracy

	LR		NN		Pure SVM		GA-SVM	
	Training	Validation	Training	Validation	Training	Validatio n	Training	Validation
32FS	78.13%	68.18%	79.57%	68.18%	82.45%	72.22%	86.53%	80.30%
30FS	80.53%	67.68%	78.85%	69.19%	84.86%	71.72%		
12FS	66.83%	68.69%	79.81%	69.19%	81.73%	72.73%		
6FS	76.92%	70.71%	75.48%	71.72%	81.01%	74.75%		

Table 5 shows the results of McNemar test. As shown in Table 5, GA-SVM outperforms LR and NN with the 1% statistical significant level and Pure SVM with the 10% statistical level. But the other models do not significantly outperform each other.

Table 5. p values for the validation data

	NN	Pure SVM	GA-SVM
LR	0.727	0.115	0.002***
NN		0.263	0.004***
Pure SVM			0.082*

* significant at the 10% level
*** significant at the 1% level

6. Conclusion

In this paper we dealt with the problem of feature selection for SVM by means of GA. Additionally to the selection of a feature subset, GA is also used to optimize parameters of SVM. The proposed model, GA-SVM, optimizes feature subset and parameters of SVM simultaneously.

We investigate to develop a hybrid prediction model of selecting an optimal value of parameters and feature set in SVM for the best prediction performance. Our experimentation results demonstrate that the choice of the feature subset has an influence on the appropriate kernel parameters and vice versa.

We evaluated the proposed model on real data set and compared it with other models. The results show the proposed model's effectiveness of finding optimal feature subset and parameters of SVM, and its improvement in predicting bankruptcy.

For future work, we intend to optimize kernel function, parameters and feature subset simultaneously. We would also like to expand this model to apply to instance selection problem.

References

[1] E.L. Altman, Financial Ratios, Discriminate Analysis and the Prediction of Corporate Bankruptcy, The Journal of Finance 23 (3) (1968) 589-609.

[2] E.L. Altman, I. Edward, R. Haldeman, P. Narayanan, ZETA Analysis: A New Model to Identify Bankruptcy Risk of Corporations, Journal of Banking and Finance 1 (1977) 29-54.

[3] W. Beaver, Financial ratios as predictors of failure, Empirical Research in Accounting: Selected studied, Journal of Accounting Research (1966) 71-111.

[4] J.M. Bishop, M.J. Bushnell, A. Usher, and Westland, Genetic optimization of neural network architectures for color recipe prediction, Artificial pleural Networks and Genetic Algorithms, (Springer-Verlag, New York, 1993) 719-725.

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

[6] S.M. Bryant, A Case-based reasoning approach to bankruptcy prediction modeling, International Journal of Intelligent Systems in Accounting, Finance and Management 6 (3) (1997) 195-214.

[7] P. Buta, Mining for financial knowledge with CBR. AI Expert 9 (10) (1994) 34-41.

[8] L. Cao, F.E.H. Tay, Financial Forecasting Using Support Vector Machines, Neural Computing & Applications 10 (2001) 184-192.

[9] J.R. Coakley, C.E. Brown, Artificial neural networks in accounting and finance: Modeling issues. International Journal of Intelligent Systems in Accounting, Finance and Management 9 (2) (2000) 119-144.

[10] C. Chang, C. Lin, LIBSVM: A Library for Support Vector Machines, 2001. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

[11] A.I. Dimitras, S.H. Zanakis, C. Zopounidis, A Survey of business failure with an emphasis on prediction methods and industrial applications, European Journal of Operational Research 90 (3) (1996) 487-513.

[12] A. Fan, M. Palaniswami, Selecting Bankruptcy Predictors Using A support Vector Machine Approach, Proceeding of the International Joint Conf. on Neural Network 6 (2000) 354-359

[13] D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning (Addison-Wesley, New York 1989)

[14] I. Han, J.S. Chandler, T.P. Liang, The impact of measurement scale and correlation structure on classification performance of inductive learning and statistical methods. Expert System with Applications 10 (2) (1996) 209-221.

[15] Zan. Huang, Hsinchun. Che, Chia-Jung Hsu, Wun-Hwa Chen, Soushan Wu, Credit rating analysis with support vector machines and neural networks: a Market comparative study, Decision Support Systems 37 (2004) 543-558.

[16] J. H. Holland, Adaptation in natural and artificial systems (The University of Michigan Press, Ann Arbor, 1975)

[17] L.B. Jack, A.K. Nadi, Support vector machine for detection and characterization of rolling element bearing faults, Proceedings of Institution of Mechanical Engineers. Part C: Journal of Mechanical Engineering Science 215(2000) 1065-1074.

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

[19] T. Joachims, Text Categorization with Support Vector Machines, Technical report, LS VIII Number 23, University of Dormund (1997)

[20] K. Kim, Financial Time Series Forecasting using Support Vector Machines, Neurocomputing 55 (2004) 307-319.

[21] K. C. Lee, I. Han, Y. Kwon, Hybrid NN models for Bankruptcy Predictions, Decision Support Systems 18 (1996) 63-72.

[22] T. Liang, J. Chandler, I. Hart, Integrating Statistical and Inductive Learning Methods for Knowledge Acquisition, Expert Systems with Applications I (1990) 391-401.

[23] P.A. Meyer, H. Pifer, Prediction of Bank Failures, The Journal of Finance 25 (1970) 853-868.

[24] M. Mitchell, An Introduction to Genetic Algorithms (MIT Press, Cambridge, MA Addison-Wesley, 1996)

[25] J. Ohlson, Financial Ratios and the Probabilistic Prediction of Bankruptcy, Journal of Accounting Research 18 (1) (1980) 109–131.

[26] C. Pantalone, M. B. Platt, Predicting Commercial Bank Failure since Deregulation, New England Economic Review (1987) 37–47.

[27] M. S. Schmidt, Identifying Speaker with Support Vector Networks, In Interface '96 Proceedings, Sydney (1996)

[28] B. Sclkopf, C. Burges, V. Vapnik, Extracting support data for a given task. In U.M Fayyad and R. Uthurusamy, editors, Proceedings, First International Conference on Knowledge Discovery & Data Mining. AAAI Press, Menlo Park, CA (1995)

[29] J. D. Shaffer, D. Whitley, L.J. Eshelman, Combination of genetic algorithms and neural networks: a survey of the state of art, Proceedings of International Workshop on Combination of Genetic Algorithms and Neural Networks, Baltimore, (June 1992) 1-37.

[30] M. Shaw, J. Gentry, Using and Expert System with Inductive Learning to Evaluate Business Loans. Financial Management 17(3) (1998) 45-56.

[31] S. Siegel, Nonparametric Statistics for the Behavioral Sciences. (McGraw-Hill, NY, 1956)

[32] Z. Sun, G. Bebis, R. Miller, Object Detection using Feature Subset Selection, Pattern Recognition 27 (2004) 2165-2176

[33] K. S. Tang, K. F. Man, S. Kwong, Q. He, Genetic Algorithms and Their Applications, IEEE Signal Processing Magazine 13 (1996) 22–37.

[34] F.E.H. Tay, L. Cao, Application of Support Vector Machines in Financial Time Series Forecasting, Omega 29 (2001) 309-317.

[35] F.E.H. Tay, L. Cao, Modified Support Vector Machines in Financial Time Series Forecasting, Neurocomputing 48 (2002) 847-861.

[36] T. Van Gestel, B. Baesens, J. Suykens, M. Espinoza, D.-E. Baestaens, J. Vanthienen, B. De Moor, Bankruptcy Prediction with Least Squares Support Vector Machine Classifiers, Computational Intelligence for Financial Engineering, 2003, Proceeding 2003. IEEEE International Conference on 2003 1-8.

[37] VN.Vapnik, The Nature of Statistical Learning Theory (New York, Springer-Verlag, 1995)

[38] G. Zhang, M.Y. Hu, B.E. Patuwo, D.C. Indro, Artificial Neural Networks in Bankruptcy Prediction: General Framework and Cross-validation Analysis, European Journal of Operational Research 116 (1) (1999) 16-32.