Asia Pacific Journal of Information Systems
Vol. 21, No. 2, June 2011

Feature Selection for Multi-Class Support Vector Machines Using an Impurity Measure of Classification Trees: An Application to the Credit Rating of S&P 500 Companies

Taeho Hong*, Jiyoung Park**

Support vector machines (SVMs), a machine learning technique, has been applied to not only binary classification problems such as bankruptcy prediction but also multi-class problems such as corporate credit ratings. However, in general, the performance of SVMs can be easily worse than the best alternative model to SVMs according to the selection of predictors, even though SVMs has the distinguishing feature of successfully classifying and predicting in a lot of dichotomous or multi-class problems. For overcoming the weakness of SVMs, this study has proposed an approach for selecting features for multi-class SVMs that utilize the impurity measures of classification trees. For the selection of the input features, we employed the C4.5 and CART algorithms, including the stepwise method of discriminant analysis, which is a well-known method for selecting features. We have built a multi-class SVMs model for credit rating using the above method and presented experimental results with data regarding S&P 500 companies.

Keywords : Information Technology, Feature Selection, Multi-class SVMs, Impurity Measures, Credit Rating

* Associate Professor, School of Business, Pusan National University
** Corresponding Author, Post-Doc, BK21 Research and Education Institute, Pusan National University

Ⅰ. Introduction

It has become increasingly important to effectively keep, manage, and make full use of data in large databases since database was introduced to manipulate data. Data mining is the process of exploration and analysis, by automatic or semi-automatic means, of large quantities of data in order to discover meaningful patterns and rules [Berry and Linoff, 1997]. That is why we adopt data mining to search useful information in large databases where there is irrelevant or redundant information as well as important information that is related to the research objectives. Machine learning algorithms have been generally used in data mining such as neural networks, support vector machines since they showed distinguished performance in a lot of data mining tasks, whichever are designed for prediction or classification. For better performance and higher quality of machine learning algorithms, we can use various approaches such as reducing the number of predictors and controlling or eliminating noise in the data-processing stage. Among these approaches, feature selection is one of the most important issues in many data mining tasks; appropriate feature selection can enable a less computationally intensive model and the best possible performance, and improve the effectiveness and domain interpretability of the inference model [Huang *et al.*, 2008; Su and Yang, 2008]. Namely, a small feature set is sufficient for enhancing the generalization and presenting the interpretability of the model. Features are akin to predictors, attributes, or properties. For example, in credit-rating problems, features might be many financial predictors. Parsimony and compactness are desirable features in credit-rating models for gaining more accurate models and avoiding some problems. There are several reasons why we should reduce the number of predictors in the models instead of using all of them [Shmueli *et al.*, 2007]: 1) it may be expensive or not feasible to collect a full complement of predictors for future predictions, 2) the more predictors there are, the higher is the chance of missing values in the data, 3) parsimony is an important property of good models, 4) it can be shown that the use of predictors that are uncorrelated with the dependent variable increases the variance of prediction, and 5) it can be shown that dropping predictors that are actually correlated with the dependent variable can increase the average error in prediction.

This study selected influential predictors to the credit ratings of companies from the set of financial features by using impurity measures from decision trees and to build a classification model using multi-class support vector machines (SVMs) for credit rating. To select the predictors for multi-class SVMs, we utilized the impurity measures of classification trees. The Gini index and entropy measure are well known and popular although there are various ways to measure impurity [Shmueli *et al.*, 2007]. C4.5 [Quinlan, 1993] and CART [Breiman *et al.*, 1984], which are algorithms for learning in classification trees, have been frequently employed in the classification domains because of their capability to generalize and explain decision trees and present the rules. Since there are some decisive features that influence the target variable in the datasets, we have to find essential features with the impurity measures to reduce the error rates or complexity of models.

To develop classification models, we employed multi-class SVMs. SVMs, a data mining technique, has shown outstanding performance and a generalizing capacity in classification problems. In addition, SVMs has attracted many researchers in the context of accuracy and explanatory capability as compared to neural networks, which are one of the most representative techniques in data mining. SVMs can be applied to not only binary classifications such as bankruptcy but also multi-class problems such as credit ratings, as in this study. The credit rating of corporations, which is the most typical example of multi-class problems, encompasses two types of approaches for solving multi-class problems: binary-based methods and all-together methods. The former type of approaches comprises one-against-all, one-against-one, and directed acyclic graphs (DAGSVMs), and the latter encompasses Weston and Watkins [1999] and Crammer and Singer [2000]. We employed one-against-one as a binary-classification based method and the Weston and Watkins [1999] and Crammer and Singer [2000] methods as an all-together based method. The classification of the credit ratings of corporations has been studied extensively since data mining was applied to predict the classes. Researchers have developed the classification models using neural networks [Dutta and Shekhar, 1998; Garabaglia, 1991; Kim, 1993; Singleton and Surkan, 1990], inductive learning and case-based reasoning [Shin and Han, 2001], and SVMs [Ahn *et al.*, 2006; Huang *et al.*, 2004; Lee, 2007].

This study proposed a method that is applicable to the classification of credit ratings in actual practice. We can easily classify companies through well-made models. In other words, by replacing human labor by data-driven methods, we can solve the classification problems more efficiently. In addition, using different features selected by the stepwise method, the Gini index, and the entropy measure, we developed multi-class SVMs models. To solve the multi-class problems, we employed one-against-one as a binary-classification based method and Weston and Watkins [1999] and Crammer and Singer [2000] methods as an all-together based method. We developed nine SVMs models using a combination of three feature selection methods and three classification methods. The comparative analyses for nine SVMs models were presented with the credit ratings of S&P 500 companies.

Ⅱ. Literature Review

2.1 Feature selection

In the feature selection, first off we have to consider several reasons why we should reduce the number of variables in the models instead of using all of them [Shmueli *et al.*, 2007]. For example, it may be expensive or in feasible to collect a full complement of predictors. Also, the more are the predictors, the more are the more missing values in most cases. Parsimony is an important property of a good model. The variance of prediction increases because of the presence of predictors that are uncorrelated with the dependent variable. Namely, small feature set has the more interpretability and the minimum amount number of predictors can enhance the generalization and enable simple rules in the model. Furthermore, we can make a good decisions using the check list of Guyon and Elisseeff [2003]: understanding of the domain knowledge, commensurable features, interdependences of features, necessity of pruning input features, necessity of assessing features individually, necessity of a predictor, consideration of noisy data, thinking about what to try first, consideration about new ideas, time, computational resources, sufficiency of examples, and the need for a stable solution to improve performance or understanding. Liu and Motoda [1998] introduced methods of the extraction of key features, as follows: 1) feature construction is a process that discovers missing information about the relationships between features and augments the space of features by inferring or creating additional features, 2) feature extraction is a process that extracts a set of new features from the original features through some functional mapping, 3) subset selection is different from feature transformation in that no new features will be generated; rather, only a subset of the original features is selected and the feature space is reduced. Feature selection algorithms can be classified as following embedded, filters, and wrappers approaches [Tan *et al.*, 2005]:

- Embedded approaches: Features selection occurs naturally as part of the data mining algorithm. Specifically, during the operation of the data mining algorithm, the algorithm itself decides which predictors to use and which to ignore. For example, decision tree includes algorithms for selecting features during the task of data mining.
- Filter approaches: Features are selected before the data mining algorithm is run, using some approach that is independent of the data mining task. For example, correlation analysis and t-test are used for ranking the priority of each feature.
- Wrapper approaches: These methods use the target data mining algorithm as a black box to find the best subset. For the wrapper approach, where evaluation consists of actually running the target data mining application, the subset evaluation function is simply the criterion normally used to measure of data mining. Genetic algorithms are the representative wrapper approaches integrated to neural networks, and SVMs models in data mining.

In all, the above is what we know about feature selection. To use SVMs models, which have been known as the most powerful in data mining for the credit rating, we can choose filter approach or wrapper approach. Genetic algorithms, the representative wrapper approach, have been regarded to select the best features helping neural networks or SVMs in data mining [Huang and Wang, 2006; Mao, 2004; Shin and Lee, 2002]. Their common drawback is, however, that they have a higher risk of overfitting than filter techniques and are very computationally intensive to build models. In other words, wrapper approaches are too complicated and time-consuming for a layman, who is not skillful at data mining, to be charge with his/her task for credit rating in practice.

Decision trees were preferred for the feature selection as the filter approach due to the capability of taking reduced feature set without expert knowledge, which was required to know wheatear the features were important or not in the model [Blum and Langley, 1997]. Cardi [1993] employed decision trees for the feature selection in the task of natural language processing while case-based learning systems were made without reliance on potentially expensive expert knowledge. Sugumaran *et al.* [2008] studied about fault diagnostics of roller bearing. This study addressed the feature selection process using decision tree and used kernel based neighborhood score multi-class support vector machines for classification. The advantages of decision tree algorithm are to identify the good features for the purpose of classification from given training data set, and to reduce domain knowledge required to select good features for the feature selection of the pattern classification problem. Lin *et al.* [2009] employed SVMs and decision tree to perform a feature selection and a parameter determination. This study was about the applying enhanced data mining approaches in predicting bank performance on a case of Taiwanese commercial banks. Experimental results showed that used methods were effective in searching for the beneficial subset of features and parameter values. In this study, we have employed inductive learning algorithms, specif., C4.5 and CART among filter approaches to develop SVMs models for credit rating with small feature sets. However the proposed model would be comparable to the best in previous models if we compared the performance of classification capability among the models. In addition, the reason why we use inductive techniques for features subset selection is that these techniques can select influential predictors easily. The inductive techniques look for the best trees with the lowest impurity through the learning of data set. A number of algorithms have been developed to measure impurity since ID3 algorithm for inductive learning was introduced by Quinlan in 1982. The two most popular measures are the Gini index and the entropy measure. To perform comparative analysis between the impurity measures of decision trees and popular methods of selecting predictors for classification model, we undertook multidiscriminant analysis (MDA) by selecting predictors in a stepwise manner.

2.2 Multi-Class Support Vector Machines

Support Vector Machines (SVMs) became one of the popular techniques for solving data-classification problems. This method has been applied to not only binary classification such as bankruptcy but also multi-class problems such as credit rating, which was considered in this study. The SVMs method has shown outstanding performance and a generalizing capacity in classification tasks since it was developed from statistical theory by Vapnik [1995]. SVMs is an algorithm that finds the maximum margin hyperplane, which is the maximum separation between classes. Here, support vectors are the closest to the maximum margin hyperplane. If it is hard to divide the data into two classes, we introduce the kernel function. In the case of nonlinear class boundaries, as the original input space is mapped into a high-dimensional dot-product space by the help of the kernel function, we transform the inputs into a high-dimensional feature space.

In the two separated classes, A is defined as $x_i \in R^n$, $y_i = 1$, while B is defined as $x_i \in R^n$, $y_i = -1$ if we presume the function, $f: R^n \rightarrow \{\pm 1\}$, by using a training set. If it is possible to separate them linearly, they can be represented in Equation (1) and Equation (2).

$$w \cdot x_i + b \geq \pm 1, \ \forall x_i \in A \tag{1}$$

$$w \cdot x_i + b \leq -1, \ \forall x_i \in B \tag{2}$$

In the above, x is the input vector, w is the weight vector, and b is the bias. w and b represent the parameters that are used to determine the hyperplane.

Using Equation (1) and Equation (2), we can derive Equation (3) as follows:

$$y_i(w \ x_i + b) \geq 1, \ \forall x_i \in A \cup B. \tag{3}$$

The maximum margin classifier optimizes data within the maximum margin hyperplane. This is an optimization problem that is expressed in Equation (4):

$$\min_{w,b} \frac{w \cdot w}{2}$$
$$s.t. \ y_i(w \cdot x_i + b) \geq 1 \tag{4}$$

Finally, the equation for an optimal separating hyperplane is shown below.

$$f(x, a_i, b) = \sum y_i \alpha_i (x \ x_i) + b. \tag{5}$$

In the above, α_i and b are parameters for determining the separation of the hyperplane. x denotes the training data, and x_i is the support vector.

In the case of nonlinear class boundaries, we can implement the idea by transforming the inputs into a high-dimensional feature space. A nonlinear case of separation is represented in Equation (6).

$$f(x, \alpha_i, b) = \sum y_i \alpha_i K(x, x_i) + b \tag{6}$$

In the above, $K(x, x_i)$ is called the kernel function, which we employed in this study. Examples of the kernel function are as follows.

- The polynomial kernel:
  $K(x, x_i) = (x \cdot x_i + 1)^d$.
- The Gaussian radial basis function (RBF):
  $K(x, x_i) = \exp(\frac{1}{\delta^2}(x \ x_i)^2)$.

Many studies have employed SVMs to make important decisions for the issues occurred in business as follows: the classification of credit ratings [Ahn *et al.*, 2006; Chen and Shih, 2006; Gestel *et al.*, 2006; Huang *et al.*, 2004; Huang *et al.*, 2008; Lee, 2007], the prediction of bankruptcy [Min and Lee, 2005; Shin *et al.*, 2005; Wu *et al.*, 2007], the classification of customer level [Kim and Ahn, 2010], efficiency rating of venture business [Park and Hong, 2009], and the forecasting of financial time series [Kim, 2003; Tay and Cao, 2001]. The credit ratings are generally over three at least and several classes should be classified, which is so-called a multi-class problem. For multi-class problems such as the classification of credit rating, binary classification-based methods and all-together methods are commonly employed to solve. In other words, there are two types of approaches for multi-class SVMs. One is by constructing and combining several binary classifiers, such as one-against-all, one-against-one, and directed acyclic graph SVMs (DAGSVMs), and the other is by directly considering all the data in an optimization formulation [Hsu and Lin, 2002]. The latter encompasses all-together methods such as Weston and Watkins [1999] and Crammer and Singer [2000]; these have different decomposition me-

<Table 1> Studies using Multi-class SVMs for Credit Ratings

Studies	Feature selection methods	Classification methods
Ahn and Kim [2009]	one-way ANOVA, stepwise MDA	OAO, OAA, DAGSVM, ECOC, W&W, C&S
Ahn *et al.* [2006]	literature study, t-test, stepwise MDA	OAO, W&W, C&S
Chen and Shih [2006]	financial analysts	OAO
Gestel *et al.* [2006]	financial analysts	OAO
Huang *et al.* [2004]	literature study, ANOVA	OAO, C&S
Lee [2007]	stepwise MDA	OAO

Note) OAO: One-against-one, OAA: One-against-all, DAGSVM: directed acyclic graph SVMs, ECOC: Error-Correcting Output Coding, W&W: Weston and Watkins, C&S: Crammer and Singer.

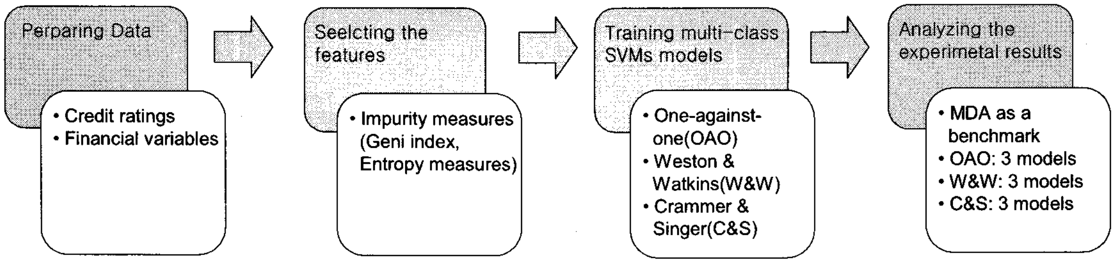
thods.

In this study, we used the Gaussian radial basis function (RBF) as a kernel function of SVMs, and adopted a one-against-one approach among binary classification methods and two all-together methods that were proposed by Weston and Watkins [1999] and Crammer and Singer [2000] each to solve multi-class problems.

Ⅲ. Research Framework

In this study, we have proposed a method to select the features of classification trees for multi-class SVMs using impurity measures: the Gini index and the entropy measure. We have applied our approach to classify the credit ratings of S&P 500 companies. Through the framework of feature selection, we can consider the importance of parsimony in developing good models. The main objectives of this study are to select the important features that are correlated with credit ratings using impurity measures among many financial features, and to develop multi-class SVMs models for credit-rating classification with the selected influential features.

Two approaches for solving multi-class problems are relevant in this paper to the credit-rating problem with regard to S&P 500 companies; this problem is the most typical example of multi-class problems. One is a binary-classification based method and the other is an all-together based method. We employed one-against-one as a binary-classification based method and the Weston and Watkins [1999] and Crammer and Singer [2000] methods as an all-together based method. In SVMs, we have to choose the noise level C, the kernel function, and its parameters. We can ascertain the parameters of the kernel function and C by the trial-and-error method. Through the feature selection method and the type of multi-class SVMs, we obtained nine models for multi-class SVMs according to the three approaches for solving multi-class problems and three methods for selecting features. To perform comparative analysis between our proposed method and a popular method of selecting features for the classification model, we used multidiscriminant analysis (MDA) by selecting features in a stepwise manner. We have proposed a model that is applicable to the classification of bond ratings in actual practice. In other words, by replacing human labor through



<Figure 1> Research Procedures

a data-driven method, we can solve the classification problems more efficiently. <Figure 1> shows our research procedures as follows.

- Step 1: Preparing experimental data that have credit ratings and financial features.
- Step 2: Selecting the features from impurity measures: the Gini index and entropy measures.
- Step 3: Training multi-class SVMs model with the selected features according to three classification methods: one-against-one, Weston and Watkins [1999], and Crammer and Singer [2000].
- Step 4: Analyzing the experimental results for proposed models.

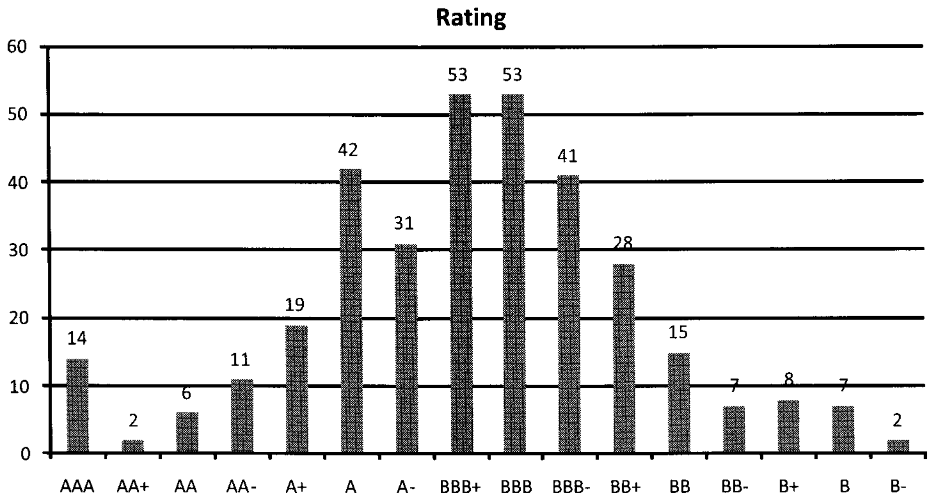
Ⅳ. Experiments and Results

4.1 Data

Our dataset consists of 339 companies listed in the S&P 500 index during 2003~2008 year, which was collected from the database of Data Stream in Thomson. Standard and Poor's credit rating sets are comprised of 10 grades ranging from 'D' to 'AAA' and intermediate ratings are ggiven to each level between AA and CCC (e.g., BBB+, BBB and BBB-). According to Standard and Poor's issuance of credit ratings, an obligation rated 'AAA' has the highest rating. The obligor's capacity to meet its financial commitment on the obligation is extremely strong. An obligation rated 'BBB' exhibits adequate protection parameters. However, adverse economic conditions or changing circumstances are more likely to lead to a weakened capacity of the obligor to meet its financial commitment on the obligation. In the dataset of this study, 'AAA' companies are about 4.1% of the total while 'BBB+', 'BBB', and 'BBB-' companies are in the greatest proportions. <Figure 2> and <Table 2> present the organization of the dataset; we can establish a numerical difference in the credit ratings. Credit ratings were transferred into new groups ranging from 1 to 6 according to the actual credit levels in our models considering the frequency of each class. The data were separated into training and testing datasets. The training dataset consists of 272 firms, and the test dataset consists of 67 firms.

4.2 Feature selection

We prepared predictors with 35 financial features, which were commonly considered in the task of credit rating before selecting features



<Figure 2> Ratings Distribution in the Actual Dataset

<Table 2> Frequency and Proportion of the Credit Ratings in the Datasets

Actual Rating	Model Rating	Frequency	%Ratio	Training	Test
AAA	1	14	4.1	11	3
AA+, AA, AA-	2	19	5.6	15	4
A+, A, A-	3	92	27.1	74	18
BBB+, BBB, BBB-	4	147	43.4	118	29
BB+, BB, BB-	5	50	14.7	40	10
B+, B, B-	6	17	5.0	14	3
Total	-	339	100	272	67

for multi-class SVMs in the proposed way. By employing the stepwise method of discriminant analysis, which is a well-known method for selecting features, the predictors for SVMs were determined for the purposed of comparing our proposed model. The features, which would be the most significant in predicting credit ratings, were selected by minimizing the impurity in the Gini index and entropy measures gained the C4.5 and CART algorithms respectively. Through the feature selection process presented in our approach, the three set of predictor were used for SVMs modeling of the credit-rating classification and they are shown in <Table 3>. Four, seven, and six predictors, respectively, were picked up by the MDA, C4.5, and CART classifiers. For this process, we utilized SAS Enterprise Miner 4.3 and SPSS 15.0. We found that the selected features differed from each other according to the methods for feature selection. In the feature-selection process, we have to consider the characteristics of a good model discussed above. For example, a small feature set has greater interpretability and the

<Table 3> Feature Selection from the MDA, C4.5, and CART Classifiers

Features	MDA	C4.5	CART
Market capitalization	●	●	●
Selling and administration expenses		●	
Income tax		●	●
Earnings before interest and taxes		●	●
Cash and short term investment		●	
Total assets		●	●
Total capital	●		
Current ratio	●		
Total debt as a % of common equity	●		
Dividends per share			●
Earnings per share		●	●

minimum number of features can enhance generalization and lead to simple rules for the model. In the results of multi-class SVMs models reported in the next section, we tried to determine which method was excellent in terms of feature selection.

4.3 SVMs Model Specification and Results

We have employed the impurity measures of classification trees for feature selection, and developed multi-class SVMs models using the selected features. To develop SVMs models, we had to choose the kernel function, and set the noise level C and kernel parameters. The upper bound C and the kernel parameter δ^2 play important roles by effecting a great change of the performance in SVMs models. We employed BSVM-2.06 for the SVMs experiment and RBF was used as the kernel function of SVMs. We simulated the following parameters for asserting the optimal SVMs model by trial and error.

- $C \in \{1, 20, 40, 60, 80, 100, 200, 1000\}$
- $\delta^2 \in \{1, 0.7, 0.5, 0.2, 0.1, 0.05, 0.01, 0.005\}$

The experimental results are presented in <Table 4>. To perform comparative analysis between our proposed method and a popular method of selecting features for the classification model, we employed MDA by selecting the features in a stepwise manner. We applied the one-against-one approach as the binary classification method and the approaches of Weston and Watkins [1999] and Crammer and Singer [2000] as the all-together methods to develop multi-class SVMs models. Finally, we obtained nine models for multi-class SVMs according to the three approaches for solving multi-class problems and three methods for feature selection. We found that the SVMs model with Crammer and Singer's [2000] method yields a classification accuracy of 53.7% when the predictors are selected by minimizing the Gini index in classification trees. Also, the SVMs model with Weston and Watkins's [1999] method results in

<Table 4> Results of the Multi-Class SVMs Models

No.	Model	SVMs Parameters		Accuracy	
		C	δ^2	TR(%)	TS(%)
1	MDA	-	-	36.0	34.3
2	OAO-SVMs-MDA	80	0.01	50.7	47.8
3	OAO-SVMs-C4.5	1000	0.2	71.0	50.7
4	OAO-SVMs-CART	100	0.7	72.1	50.7
5	W&W-SVMs-MDA	200	0.005	52.9	47.8
6	W&W-SVMs-C4.5	60	0.7	71.0	53.7
7	W&W-SVMs-CART	40	0.7	68.8	53.7
8	C&S-SVMs-MDA	80	0.005	50.7	47.8
9	C&S-SVMs-C4.5	60	0.7	73.9	52.2
10	C&S-SVMs-CART	40	0.7	71.7	53.7

Note) Legend: MDA; multi-discriminant analysis as a benchmark/OAO; one-against-one method/ W&W; Weston and Watkins [1999]/C&S-Crammer and Singer [2000]/OAO-SVMs-MDA; multi-class SVMs model using the one-against-one classification method and the features selected by MDA.

classification with an accuracy of 53.7% when the predictors are selected by minimizing the entropy measure in classification trees. Therefore, we concluded that our proposed method is valid for selecting the features of multi-class SVMs in credit rating. The results of the multi-class SVMs models are better than those of the stepwise method of MDA, which is a well-known method for selecting features.

We can ascertain that the performance in multi-class problems is not high enough. To solve multi-class problems, we have to consider the scarcity of data, because the number of samples is not uniform in every class. Some classes have very small samples, while some other classes do not. For this reason, the performance of the predictive model takes a sudden turn for the worse; so, we cannot validate the statistical compatibility of models in this case. <Table 5> shows the within-1-class accuracy results of the multi-class SVMs models. The range of the probabilities for the predictions was from 88.1% to 91.0% for all 9 models using the multi-class SVMs.

V. Conclusions

In this study, we proposed a method to select the features for multi-class SVMs using the impurity measures of classification trees, and applied our approach to classify the credit ratings of S&P 500 companies. The main objectives of this study are to select important predictors that are correlated with the credit rating, being the dependent variable, by using impurity measures among various financial features, and to develop multi-class SVMs models for the credit-rating classification with the selected influential features. In the feature-selection process, we employed the Gini index and entropy measure, which are the two most popular measures among many ways of measuring the impurity, and we used a stepwise method

<Table 5> Within-1-class Accuracy Results

Actual Rating	OAO-SVMs-MDA						OAO-SVMs-C4.5						OAO-SVMs-CART						Total
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
1	0	1	1	1	0	0	0	1	1	1	0	0	0	1	1	1	0	0	3
2	0	1	1	2	0	0	0	1	1	2	0	0	0	1	1	2	0	0	4
3	0	0	2	16	0	0	0	0	2	16	0	0	0	0	2	16	0	0	18
4	0	1	1	27	0	0	0	0	3	26	0	0	0	0	3	26	0	0	29
5	0	0	0	8	2	0	0	0	0	10	0	0	0	0	0	10	0	0	10
6	0	0	0	2	1	0	0	0	0	3	0	0	0	0	0	3	0	0	3
Total	0	3	5	56	3	0	0	2	7	58	0	0	0	2	7	58	0	0	67
probability	88.1%						89.6%						89.6%						

Actual Rating	W&W-SVMs-MDA						W&W-SVMs-C4.5						W&W-SVMs-CART						Total
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
1	0	1	1	1	0	0	1	1	0	1	0	0	1	1	0	1	0	0	3
2	0	1	1	2	0	0	0	2	0	2	0	0	0	2	0	2	0	0	4
3	0	0	2	16	0	2	0	0	8	10	0	0	0	0	9	9	0	0	18
4	0	0	4	22	1	0	0	1	4	21	3	0	0	1	4	20	4	0	29
5	0	0	0	8	2	0	1	0	0	3	4	2	1	0	1	2	4	2	10
6	0	0	0	2	1	0	0	0	1	1	1	0	0	0	0	1	2	0	3
Total	0	2	8	51	4	2	2	4	13	38	8	2	2	4	14	35	10	2	67
probability	88.1%						89.6%						89.6%						

Actual Rating	C&S-SVMs-MDA						C&S-SVMs-C4.5						C&S-SVMs-CART						Total
	1	2	3	4	5	6	1	2	3	4	5	6	1	2	3	4	5	6	
1	1	1	0	1	0	0	2	0	0	1	0	0	1	1	0	1	0	0	3
2	0	1	1	2	0	0	0	2	0	2	0	0	0	2	0	2	0	0	4
3	0	0	3	14	1	0	0	0	8	10	0	0	0	0	8	9	1	0	18
4	0	0	4	24	1	0	0	1	5	19	3	1	0	0	6	21	2	0	29
5	0	0	0	8	2	0	1	0	1	5	1	2	1	0	1	2	4	2	10
6	0	0	0	2	1	0	0	0	0	1	2	0	0	0	0	2	1	0	3
Total	1	2	8	51	5	0	3	3	14	38	6	3	2	3	15	37	8	2	67
probability	91.0%						88.1%						88.1%						

of discriminant analysis as a benchmark, which is a well-known method for selecting features. Experimental results showed that the three methods resulted in different predictors. Four, seven, and six predictors, respectively, were selected by the stepwise method of MDA, and the C4.5 and CART methods.

To develop multi-class SVMs models, we employed three classification methods: one-against-one, Weston and Watkins [1999], and Crammer and Singer [2000]. To perform comparative analysis between our proposed method and a popular method of selecting features for the classification model, we employed MDA by selecting the features in a stepwise manner. We built nine models for multi-class SVMs according to three approaches for solving multi-class problems and three methods for selecting features; MDA has been used as the benchmark model. From the results for multi-class SVMs models, we could determine which method was excellent in feature-selection approaches. We found that the SVMs model with the Weston and Watkins [1999] method yields a classification accuracy of 53.7% when the predictors are selected by minimizing the entropy measure in classification trees. In the results for within-1-class accuracy of the multi-class SVMs models, the hit ratio ranged from 88.1% to 91.0%. The results of multi-class SVMs models are better than those of the stepwise method of discriminant analysis, which is a well-known method for classification and feature selection. Hence, we have concluded that our proposed method is valid for selecting the features of multi-class SVMs models.

Feature selection is the foremost step that has more influence on performance in data mining than other aspects, and there are many checklists to be considered in feature selection. In various data-mining tasks, feature selection is one of the most important issues, because the appropriate features can result in excellent models with good interpretability, clarity, and conciseness. When we develop SVMs models with many features, from the viewpoint of data parsimony, we have to consider many aspects about feature selection such as domain knowledge, various problems about features, a stable solution, and so on. Our approach helps data miners to select or reduce the number of features in models, and solves classification problems in actual practice in many domains including the financial domain. In other words, by replacing human labor by a data-driven method, we can solve classification problems more efficiently.

〈References〉

[1] Ahn, H. and Kim, K.-J., "Corporate Bond Rating using Various Multiclass Support Vector Machines," *Asia Pacific Journal of Information Systems*, Vol. 19, No. 2, 2009, pp. 157-178.

[2] Ahn, H., Kim, K.-J., and Han, I., "Intelligent Credit Rating Model for Korean Companies using Multiclass Support Vector Machines," *Korean Management Review*, Vol. 35, No. 5, 2006, pp. 1479-1496.

[3] Berry, M.J.A. and Linoff, G.S., *Data Mining Techniques*, New York: Wiley, 1997.

[4] Blum, A.L. and Langley, P., "Selection of Relevant Features and Examples in Machine Learning," *Artificial Intelligence*, Vol. 97, No. 1-2, 1997, pp. 245-271.

[5] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*. Wadsworth International Group, 1984.

[6] Cardie, C., "Using Decision Trees to Improve Case-Based Learning," *Proceedings of the Tenth International Conference on Machine Learning*, Morgan Kaufmann, 1993, pp. 25-32.

[7] Chang, C.-C. and Lin, C.-J., BSVM: http://www.csie.ntu.edu.tw/~cjlin/bsvm/, 2006.

[8] Chen, W.-H. and Shih, J.-Y., "A study of Tiwan's issuer credit rating systems using support vector machines," *Expert Systems with Applications*, Vol. 39, No. 3, 2006, pp. 427-435.

[9] Crammer, K. and Singer, Y., "On the learnability and design of output codes for multiclass problem," *In proceedings of the Thirteenth Annual Conference on Computational Learning Theory*, 2000, pp. 35-46.

[10] Dutta, S. and Shekhar, S., "Bond rating: a non-conservative application of neural networks," *In Proceedings of IEEE International Conference on Neural Networks*, 1998, pp. II443-II450.

[11] Garabaglia, S., "An application of a Counter-Propagation Neural Networks: Simulating the Standard and Poor's Corporate Bond Rating Systems," *In Proceedings of the First International Conference on Artificial Intelligence on Wall Street*, 1991, pp. 278-287.

[12] Gestel, T.V., Baesens, B., Dijcje, P.V., Garcia, J., Suykens, J.A.K., and Vanthienen, J., "A Process model to develop an internal rating system: sovereign credit ratings," *Decision Support Systems*, Vol. 42, No. 2, 2006, pp. 1131-1151.

[13] Guyong, I. and Elisseeff, A., "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, Vol. 3, 2003, pp. 1157-1182.

[14] Hsu, C.W. and Lin, C.J., "A Comparison of Methods for Multiclass Support Vector Machines," *IEEE Transactions on Neural Networks*, Vol. 13, No. 2, 2002, pp. 415-425.

[15] Huang, C.-L., Liao, H.-C., and Chen, M.-C., "Prediction model building and feature selection with support vector machines in breast cancer diagnosis," *Expert Systems with Applications*, Vol. 34, No. 1, 2008, pp. 578-587.

[16] Huang, C. and Wang, C., "A GA-based feature selection and parameters optimization for support vector machines," *Expert Systems with Applications*, Vol. 31, No. 2, 2006, pp. 231-240.

[17] Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., and Wu, S., "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decision Support Systems*, Vol. 37, No. 4, 2004, pp. 543-558.

[18] Park, J. and Hong, T., "The prediction of DEA based efficiency rating for venture business using multi-class SVM," *Asia Pacific Journal of Information Systems*, Vol. 19, No. 2, 2009, pp. 139-155.

[19] Kim, J.W., "Expert systems for bond rating: a comparative analysis of statistical, rule-based and neural network systems," *Expert Systems*, Vol. 10, No. 3, 1993, pp. 167-171.

[20] Kim, K.-J., "Financial time series forecasting using support vector machines," *Neurocomputing*, Vol. 55, No. 1-2, 2003, pp. 307-319.

[21] Kim, K. and Ahn, H., "Customer level classification model using ordinal multiclass support vector machines," *Asia Pacific Journal of Information Systems*, Vol. 20, No. 2, 2010, pp. 21-37.

[22] Lee, Y.-C., "Application of support vector machines to corporate credit rating prediction," *Expert Systems with Applications*, Vol. 33, No. 1, 2007, pp. 67-74.

[23] Lin, S.-W., Shiue, Y.-R., Chen, S.-C., and Cheng, H.-M., "Applying enhanced data mining approaches in predicting bank performance: A case of Taiwanese commercial

banks," *Expert Systems with Applications*, Vol. 36, No. 9, 2009, pp. 11543-11551.

[24] Liu, H. and Motoda. H. *Feature Extraction, Construction and Selection: A Data Mining Perspective*, Kluwer Academic Publishers, 1998.

[25] Mao, K.Z., "Feature subset selection for support vector machines through discriminative function pruning analysis," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 34, No. 1, 2004, pp. 60-67.

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

[27] Quinlan, J.R., C4.5: *Programs for Machine Learning*, Los Altos, California: Morgan Kaufmann Publishers, Inc., 1993.

[28] Shin, K.-S. and Han, I., "A case-based approach using inductive indexing for corporate bond rating," *Decision Support Systems*, Vol. 32, No. 1, 2001, pp. 41-52.

[29] Shin, K.S. and Lee, Y.J., "A genetic algorithm application in bankruptcy prediction modeling," *Expert Systems with Applications*, Vol. 23, No. 3, 2002, pp. 321-328.

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

[31] Shmueli, G., Patel, N.R., and Bruce, P.C. *Data Mining for Business Intelligence: Concepts, Techniques, and Applications* in Microsoft Office Excel with XLMiner, Wiley InterScience, 2007.

[32] Singleton, J.C. and Surkan, A.J., "Neural networks for bond rating improved by multiple hidden layers," *In Proceedings of IEEE International Conference on Neural Networks*, 1990, pp. 163-168.

[33] Su, C.-T. and Yang, C.-H., "Feature selection for the SVM: An application to hypertension diagnosis," *Expert Systems with Applications*, Vol. 34, No. 1, 2008, pp. 754-763.

[34] Sugumara, V., Sabareesh, G.R., and Ramachandran, K.I., "Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine," *Expert Systems with Applications*, Vol. 34, No. 4, 2008, pp. 3090-3098.

[35] Tan, P., Steinbach, M., and Kumar, V. *Introduction to Data Mining*, Addison Wesley, 2006.

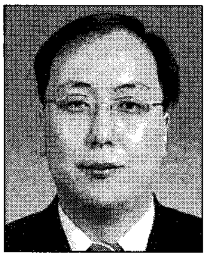
[36] Tay, F.E.H. and Cao, L., "Application of support vector machines in financial time series forecasting," *Omega*, Vol. 29, 2001, pp. 309-317.

[37] Vapnik, V. *The nature of statistical learning theory*, New York: Springer-Verlag, 1995.

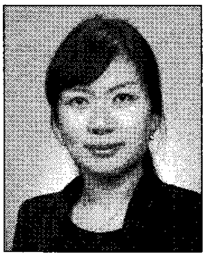
[38] Weston, J. and Watkins, C., *Multi-class support vector machines*, Presented at the Proc. ESAMM99, M. Verleysen, Ed., Brussels, Belgium, 1999.

[39] Wu, C.H., Tzeng, G.H., Goo, Y.J., and Fang, W.C., "A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy," *Expert Systems with Application*, Vol. 32, No. 2, 2007, pp. 397-408.

◆ About the Authors ◆



Taeho Hong
Taeho Hong is an associate professor in the School of Business, Pusan National University, Korea. He received a Ph.D. from the Korea Advanced Institute of Science and Technology. His research interests include data mining, CRM, and Business Intelligence. He has published in Expert Systems, Expert Systems with Applications, the Journal of MIS Research, Journal of Intelligence and Information Systems, and the Journal of Information Systems.



Jiyoung Park
Jiyoung Park is a post-doctoral researcher at the BK21 Research and Education Institute of Banking, Securities, and Derivatives for Developing Financial Hub in the North-East Asia, Pusan National University. She received her Ph.D. and M.S. in Management Information Systems, and B.S. in Statistics from the Pusan National University in Korea. Her research interests include data mining, CRM, Knowledge management, venture business, and social networks. She has published in academic journals including Asia Pacific Journal of Information Systems, the Journal of Information Systems, Knowledge Management Research.

Submitted : September 9, 2010 Accepted : April 27, 2011
1st revision : November 19, 2010