Asia Pacific Journal of Information Systems
Vol. 25 No. 3 (SEPTEMBER 2015), 579-596

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
http://dx.doi.org/10.14329/apjis.2015.25.3.579

Multi-Class SVM+MTL for the Prediction of Corporate Credit Rating with Structured Data

Gang Ren[a], Taeho Hong[b,*], YoungKi Park[c]

[a] *Doctoral Candidate, Department of Business Administration, Pusan National University, Korea*
[b] *Professor, Department of Business Administration, Pusan National University, Korea*
[c] *Assistant Professor, Department of Information Systems and Technology Management, George Washington University, USA*

ABSTRACT

Many studies have focused on the prediction of corporate credit rating using various data mining techniques. One of the most frequently used algorithms is support vector machines (SVM), and recently, novel techniques such as SVM+ and SVM+MTL have emerged. This paper intends to show the applicability of such new techniques to multi-classification and corporate credit rating and compare them with conventional SVM regarding prediction performance. We solve multi-class SVM+ and SVM+MTL problems by constructing several binary classifiers. Furthermore, to demonstrate the robustness and outstanding performance of SVM+MTL algorithm over other techniques, we utilized four typical multi-class processing methods in our experiments. The results show that SVM+MTL outperforms both conventional SVM and novel SVM+ in predicting corporate credit rating. This study contributes to the literature by showing the applicability of new techniques such as SVM+ and SVM+MTL and the outperformance of SVM+MTL over conventional techniques. Thus, this study enriches solving techniques for addressing multi-class problems such as corporate credit rating prediction.

Keywords: Support Vector Machines, SVM+, SVM+MTL (Multi-Task Learning), Credit Rating, Multi-Classification

Ⅰ. Introduction

Corporate bond rating is often used to assess the corporate debt, solvency, bankruptcy probability, and it is regularly published by professional rating agencies such as Standard and Poor's (S&P), Moody's Investor Service, and Fitch Ratings. Analysts of the agencies rate companies based on their analysis mostly of financial ratio data. The rating results are disseminated to customers and the extant and potential partners, and have a great impact on a corporate image and reputation.

However, service provided by professional rating agencies accompanies very high costs. More importantly, such costly services do not always reflect the default risk in real time, and the ratings are often affected by analysts' subjective opinions about a corporate to some degree. Therefore, it is crucial to develop an objective, credible prediction model of corporate credit rating (Cao et al., 2006).

This work was supported by the Financial Supporting Project of Long-term Overseas Dispatch of PNU's Tenure-track Faculty.
*Corresponding Author. E-mail: hongth@pusan.ac.kr Tel: 82515102531

It is known that, compared to the ratings from professional rating agencies, a data mining prediction model based on financial variables is more convenient to apply, less time-consuming and less costly in practice (Kim and Ahn, 2012). Thus, many studies have focused on developing credit rating prediction models using various data mining techniques, such as multiple discriminant analysis, logistic regression analysis, probit analysis, decision tree, cluster analysis, case-based reasoning, neural networks and support vector machine (SVM), to name a few. Among them, SVM has become one of the most popular techniques due to its high generalizability and explanation power.

Recently, novel advanced techniques such as SVM+ and SVM+MTL (multi-task learning) have emerged, and thus added more alternative data mining techniques to filed application, eventually opening new research fields in machine learning. For example, Vapnik and Vashist (2009) applied it in engineering field and Liang et al. (2009) employed SVM+ and SVM+MTL to some biomedical data sets from UCI machine learning repository. Especially, Ribeiro et al. (2010), Ribeiro et al. (2012) firstly applied the new techniques to the bankruptcy prediction field and showed that SVM+ technique outperformed the conventional SVM. These studies demonstrated that SVM+ outperformed SVM. In Liang et al. (2009), the prediction accuracy from high to low was: SVM+MTL, SVM+, and SVM. On the other hand, Ribeiro et al. (2010), Ribeiro et al. (2012) showed that SVM+ outperformed both SVM and SVM+MTL with an application to financial distress. However, little research has been conducted in applying SVM+ and SVM+MTL to multi-classification problems especially in the corporate credit rating prediction field.

This study aims to apply SVM+ and SVM+MTL techniques to a multi-class classification problem in the corporate credit rating. In addition, we intend to compare SVM, SVM+ and SVM+MTL regarding prediction performance by using empirical data from Korea bond rating market.

The main contribution of this article is to show the applicability of SVM+ and SVM+MTL to multi-class classification problems in the corporate credit rating. This study is to demonstrate SVM+MTL technique is outstanding compared to most frequently-used popular technique SVM and even compared to the novel technique SVM+. To achieve this objective, we adopt four multi-class approaches that have been widely used to solve multi-class SVM problems in prior studies, such as One-Against-All (OAA), One-Against-One (OAO), Directed Acyclic Graph (DAG), and Error Correcting Output Codes (ECOC). In this paper, they are realized by constructing several binary classifiers. The detailed explanation of these four approaches is presented in the next section. Further, we suggest an optimal model regarding the prediction performance by comparing SVM+ and SVM+MTL techniques with conventional SVM.

The rest of this paper proceeds as follows. Section 2 explains typical data mining techniques for credit rating and multi-class processing methods, including SVM, SVM+ and SVM+MTL, and then explains extant studies on credit rating and multi-class processing methods that use diverse techniques. Section 3 states the research framework and elaborates the advancement of our proposed model. Section 4 illustrates the experimental design, interprets the experimental results, and validates the robustness of new techniques (SVM+ and SVM+MTL). Finally, section 5 presents the conclusion of our work and some limitations of the study, and suggests the future work.

Ⅱ. Literature Review

2.1. Support Vector Machine (SVM)

Vapnik (1995) proposed SVM theory for binary classification based on the principle of structural risk minimization (SRM). The main idea of this algorithm is to find an optimal separable hyperplane that results in a maximum margin while ensuring the accuracy of classification and the good performance of generalization. For the nonlinear case, it achieves classification by mapping the input vectors into a high-dimensional feature space and then constructing a linear separable hyperplane.

Given a training set, $\{x_i, y_i\}_i^l = 1,\ x_i = \{x_i^1, x_i^2, \cdots, x_i^n\} \epsilon R^n,\ y_i = \{-1, 1\}$, where x denotes the input vectors, y denotes the output vectors, *l* represents the number of instances, and *n* represents the number of features for each instance. Simultaneously, it meets the following condition,

$$Y = \begin{Bmatrix} +1, \text{ if } w \cdot x_i + b \geq 1 - \xi_i \\ -1, \text{ if } w \cdot x_i + b \leq -1 + \xi_i \end{Bmatrix}$$

where w represents the complexity of models (margin width). SVM aims to find a separating hyperplane $y_i = w \cdot \Phi(x_i) + b$ between two classes in a way to maximize a margin. This maximum problem can be transformed into the following minimum problem.

$$\min_{w} \frac{\| w \|^2}{2} + C \sum_{i=1}^{l} \xi_i$$

Subject to

$$y_i(w \cdot \Phi(x_i) + b) \geq 1 - \xi_i, \qquad i = 1, 2, \cdots, N$$

Where w represents the complexity of models (margin width), C controls the trade-off between complexity (margin width) and proportion of training errors (empirical risk).

Then, this primal problem can be transformed into the following dual form by introducing Lagrange multipliers.

$$\max_{\alpha} \sum_{i=1}^{l} \alpha_1 - \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$

The optimization problem can be solved through Karuch-Kuhn-Tucher (KKT) condition. SVM solves this high dimensional problem by employing kernel function. There are four types of kernel functions in common use: linear kernel, polynomial kernel, radial basis function (RBF) kernel, and the sigmoid kernel. One of the most commonly used kernel functions is RBF kernel function in studies. They are formulated as follows:

$$K(X_i, X_j) = \begin{cases} X_i \cdot X_j & \textit{Linear} \\ (\gamma X_i \cdot X_j + C)^d & \textit{Polynomial} \\ e - \dfrac{\| X_i - X_j \|^2}{2\delta^2} & \textit{RBF} \\ \tanh(\gamma X_i \cdot X_j + C) & \textit{Sigmoid} \end{cases}$$

Then, this optimal decision function is estimated by solving the above optimization problem:

$$y = sign(w \cdot \Phi(X) + b)$$

i.e.,

$$f(X, \alpha) = sign(\sum_{i=1}^{l} \alpha_i y_i K(X_i, X) + b$$

Where X represents the input vectors; y_i represents the target labels; l represents the number of instances; X_i denotes the support vectors.

2.2. SVM+

Vapnik (2006) proposed an SVM-based optimization formulation called SVM+ for LWSD (learning with structured data) or LUPI (learning using privileged information) formulation, which exploits group information hidden in the training data. Such group information is common in many applications with heterogeneous data.

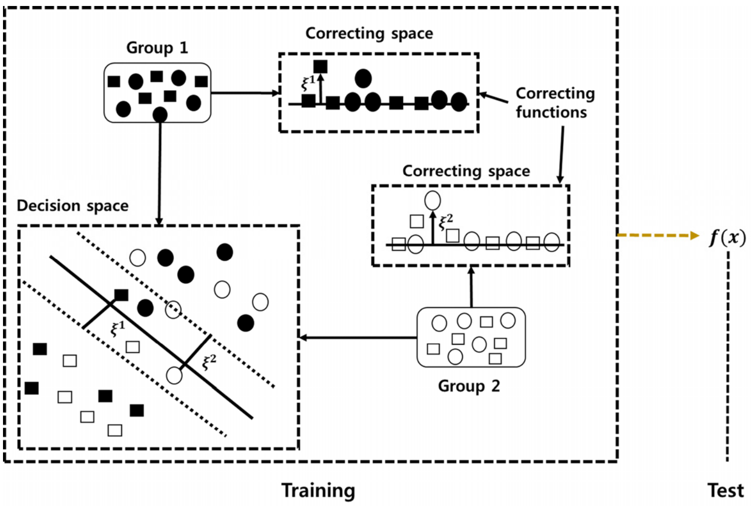
Group information (also known as grouped data, structured data, or additional information) acts as a basis for dividing data into several meaningful groups. It is used to introduce additional constraints on the slack variables (i.e., errors) for samples from different groups. For example, the corporates can be grouped by size (e.g., big, medium, small), sales, annual turnover rate, etc.

Due to the important role of group information, these grouping standards need to be selected carefully. To account for group information, Vapnik (2006) proposed a way to define the slack variables within each group by the so-called 'correcting function' and map the input vectors simultaneously into two different Hilbert spaces, as shown in <Figure 1> exhibiting for a two-group case.

Compared to standard SVM, slack variables are restricted by the correcting functions in SVM+. Mapping samples in the correcting space have to lie on one side of the corresponding function. However, slack variables are not used to assign a sample with a group membership. The data of different groups are mapped into the same decision space. But although there are different correcting functions for different groups, the correcting functions can be defined either in the same correcting space or different correcting spaces.

SVM+ maps data simultaneously into the decision space and the correcting space. Decision function is defined in the decision space, while slack correcting function is defined in the correcting space.

$$\min_{W,b\in R,\xi^r} \Phi(W,b,\xi^r) = \frac{\parallel W \parallel^2}{2} + \frac{\gamma}{2} \parallel w_R \parallel^2 + C\sum_{r=1}^{t}\sum_{i\epsilon T_r}^{l}\xi_i^r$$



<Figure 1> SVM+ Process (adapted from Liang et al. (2009))

Subject to:

$$y_i(W^T\phi(X_i)+b) \geq 1-\xi^r, \quad i \in T_{r, r=1,\cdots,t}$$
$$\xi^{r \geq 0,} i \in T_r, \quad r=1,\cdots,t$$
$$\xi^r = (W^T\phi_r(X_i)+d_r), i \in T_r, \quad r=1,\cdots,t$$

Where r denotes the number of groups; γ tunes the weight between decision space capacity and correcting space capacity; ξ^r represents the slack variables for each group and varies with r. The penalty factor C controls the trade-off between complexity and proportion of sample errors. Taken separately, SVM+ model can be divided into two parts to study. W represents the capacity of decision space, while W_r represents the capacity of correcting space. However, W_r does not decide the size of margin. Unlike conventional SVM, the term $\frac{\gamma}{2} \| W_r \|^2$ is added in SVM+ model.

Then the minimization problem can be transformed into the following dual Lagrangian form:

$$\max_{\alpha,\beta} W(\alpha,\beta) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\alpha_i\alpha_j y_i y_j K(X_i, X_j)$$
$$-\frac{1}{2\gamma}\sum_{r=1}^{t}\sum_{j\epsilon T_r}(\alpha_i+\beta_j-C)K^r(X_i^r, X_j^r)$$

Subject to:

$$\sum_{i}^{l}\alpha_i y_i = 0$$
$$\sum_{i,j\epsilon T_r}(\alpha_i+\beta_i-C)=0, \; r=1,\cdots,t$$
$$\alpha_i \geq 0, \quad \beta_i \geq 0, \quad i=1,\cdots,l$$

With the optimal that can be found by quadratic programming techniques, the form is substituted with the following final decision function to predict.

The above derivation process is similar to SVM. There are several remarks on SVM+ algorithms:

(1) The partitioning of training data into groups requires application domain knowledge or common sense, and cannot be performed using statistical analysis alone (Liang et al., 2009). Therefore, the proper selection of group variables plays a very important role.
(2) The relative performance of these methods can be strongly affected by the sample size. For a small sample size, standard SVM may still be the best method, simply because it has the fewer tunable parameters (Liang et al., 2009).
(3) The combination of kernel functions and even the combination of parameters for each method can have a significant impact on the results.
(4) If there is no structure (i.e., each training vector belongs to its own group) or there is no correlation inside groups, then the SVM+ coincides with conventional SVM (Vapnik, 2006).
(5) The drawback is that SVM+ needs more computing time than SVM (Vapnik and Vashnist, 2009).

2.3. SVM+MTL (Multi-Task Learning)

Liang et al. (2009) integrated SVM+ algorithm with multi-task learning for the reasons that (1) SVM+ can define decision functions for different groups; (2) SVM+ can model task relatedness between groups. Multi-Task Learning (MTL) is an approach that learns a problem together with other related problems at the same time, using a shared representation. The main process is briefly presented in <Figure 2>. Similar to SVM+, SVM+MTL also maps the data simultaneously into two different Hilbert spaces.

$$\min_{W, b \epsilon R, \xi^r} \Phi(W, b, \xi^r) = \frac{\| W \|^2}{2} + \frac{\gamma}{2} \| W_r \|^2 + C \sum_{r=1}^{t} \sum_{i \epsilon T_r}^{N} \xi_i^r$$

Subject to

$$y_i^r (W^T \phi_z (X_i) + b + W_r^T \phi_{(z_r)})(X_i) + d_r) \geq 1 - \xi_i^r, \quad i \in T_r, r = 1, \cdots, t$$

$$\xi_i^r \geq 0, \quad i \in T_r, \quad r = 1, \cdots, t$$

The final decision function is defined as follows.

$$f_r(x) = W^T \phi_z (X_i) + b + W_r^T \phi_{(z_r)})(X_i) + d_r r = 1, \cdots, t$$

i.e.,

$$f_r(X, \alpha) = sign(\sum_{i=1}^{l} \alpha_i y_i \phi_z (X_i, X) + b + \frac{1}{\gamma} \alpha_i y_i \phi_{z_r}(X_i) + d_r)$$

The main differences between SVM+MTL and SVM+ are summarized as follows (Liang et al., 2009):

The group information also exists in the test data for SVM+MTL. Hence, there exists one corresponding prediction model according to each group. By contrast there is only one prediction model for all groups in SVM+.

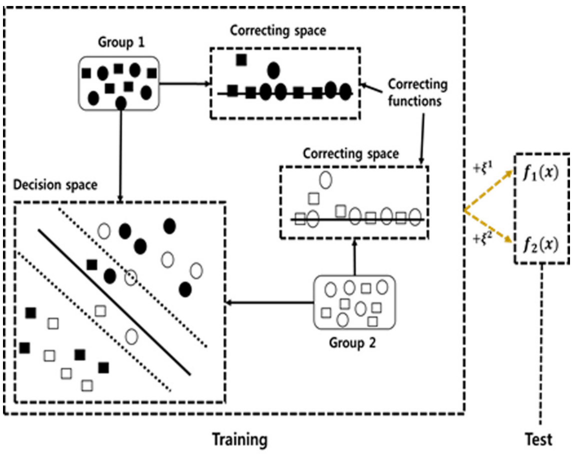
Slack variable is no longer defined by a correcting function and it represents the error of a holistic model after adding the correcting function.

Final decision function has the correcting term, which is not concluded in SVM and SVM+.

2.4. Corporate Credit Rating

Corporate credit rating, which is published by rating agencies, is an evaluation of corporate debt, solvency, and probability of default or bankruptcy. It is a reflection of potential credit risk. This section summarizes the prior studies of credit ratings that are based on data mining techniques.

Extensive studies have been done in corporate credit rating with many data mining techniques. Pinches and Mingo (1973), Pinches and Mingo (1975) utilized multiple discriminant analysis (MDA) method to develop a prediction model for industrial bond rating. Altman and Katz (1976) applied multiple logistic regression and multiple probit regression analysis to a bond rating prediction. Later, various diverse kinds of techniques are used in many studies. Here,



<Figure 2> SVM+MTL Process (adapted from Liang et al. (2009))

we focus on the summarization of credit rating studies based on SVM.

SVM algorithm has become the research focus among diverse data mining techniques. Huang et al. (2004) first employed support vector machine (SVM) in credit ratings classification to make a comparison with the neural network method. They adopted the One-Against-One and the method proposed by Crammer and Singer (2000) and demonstrated that SVM outperformed BPN and other statistical methods. Later, Chen and Shih (2006) adopted One-Against-One SVM approach for Taiwan companies and they also validated the superiority of SVM methods to logistic regression analysis and BPN methods. Cao et al. (2006) applied One-Against-All, One-Against-One, and DAGSVM (directed acyclic graph SVM) to U.S. bond rating and found that DAGSVM approach outperforms the other two approaches. Also, they demonstrated the superiority of SVM to BPN algorithm, logistic regression analysis and an ordered probit regression method. RBF kernel function and a grid-search strategy for parameters set (C, σ^2) were employed in the study in order to seek the optimal prediction value. Similar to Chen and Shih (2006), Lee (2007) also validated the superiority of SVM methods to other methods such as MDA, CBR and BPN methods in Korea market. To increase the explanatory power and stability, the study used a wider range of grid-search technique using 5-fold cross-validation to seek the optimal parameter set (C, σ^2) of RBF kernel function. Kim and Ahn (2012) nearly applied all the multi-class methods for SVM to Korea bond rating. First, they applied a method that constructs several binary classifiers, including One-Against-All, One-Against-One, DAGSVM, and ECOC (Error Correcting Output Code). Then, they applied the methods proposed by Weston and Watkins (1999) and Crammer and Singer (2000) separately, which are both based on only one classifier. Finally, they employed the OPP (Ordinal Pairwise Partitioning) method as the main methodology. All these studies are related with SVM algorithm that adopts a RBF kernel function. Further, to moderate over-fitting or under-fitting caused by the improper selection of parameters (Cao, 2006), most of these studies applied a grid search. <Table

<Table 1> Studies on Credit Rating Using Data Mining Techniques

Data Mining Techniques		Related Studies
Statistical techniques	OLS	Fisher, 1959; Horrigan, 1966; West, 1970
	MDA	Hsu and Hung, 2009; Hong and Park, 2011; Lee et al., 2002; Pinches and Mingo, 1973; Pinches and Mingo, 1975
	Logit	Bellotti et al., 2011; Kamstra et al., 2001
		Altman and Katz, 1976; Cao et al., 2006; Kamstra et al., 2001; Kim and Ahn, 2012
	Probit	Bennell et al., 2006; Bellotti et al., 2011; Cao et al., 2006
		Altman and Katz, 1976
Artificial techniques	CBR	Shin and Han, 2001; Kim and Han, 2001
	ANN	Bennell et al., 2006; Dutta and Shekhar, 1988; Huang et al., 2004; Huang et al., 2007; Kim et al., 1993; Kumar and Bhattacharya, 2006; Kwon et al., 1997; Masher and Sen, 1997; Yu et al., 2008
	SVM	Huang et al., 2004; Cao et al., 2006; Chen and Shih, 2006; Lee, 2007; Ahn and Kim, 2009; Bellotti et al., 2009; Park and Hong, 2009; Hong and Park, 2011; Kim and Ahn, 2012

1> shows the studies on credit rating using data mining techniques.

2.5. Multi-Classification Methods

Since both SVM+ and SVM+MTL techniques are all originally designed for binary classification, we have to consider the following measures that are widely used for solving multi-class SVM problems. Multi-class SVM problems are usually solved with the construction of several binary classifiers. There are four typical methods to construct binary classifiers : One-Against-All (OAA), One-Against-One (OAO), Directed Acyclic Graph (DAG), and Error Correcting Output Codes (ECOC). Hsu & Lin (2002) made a comparison of OAA, OAO, and DAG for multi-class SVM. In this study, we ran our SVM+ and SVM+MTL with the same approach. More detailed explanation of the four methods is summarized as follows.

2.5.1. One-Against-All

OAA (one-against-all) is also known as one-versus-all, one-against-rest and one-versus-rest. The separation is realized by constructing k binary classifiers in total and labeling the objective training data sets as +1 (positive) and the remaining training data sets as -1 (negative). So, the test data is classified as the class with the largest values and with the maximum distance to the positive hyperplane.

2.5.2. One-Against-One

OAO (one-against-one) method is also known as one-versus-one and pairwise classification. In this method, for a dataset with k classes (k>2), k*(k-1)/2 SVM binary classifiers are constructed to maximize the margin between any two of these classes. For the test data set, the voting strategy (a.k.a., "max-wins") is used. The corresponding class (i.e., either positive or negative class) can get one vote according to the output value (+1/-1) for each instance. Finally, the test instance is classified into the class with the largest votes, and it should be determined by the farthest hyperplane additionally when met with the tie (i.e., the same votes).

2.5.3. Directed Acyclic Graph

As the name itself implies, it is a directed graph that starts at the top node and deliveries from one node to another with no loop back to the start point eventually (Thulasiraman and Swamy, 1992). Similar to OAO method, there are $k \cdot (k-1)/2$ classifiers made by combing any two of the total classes. Yet, it is significantly different from the OAO method due to the unique directed acyclic graph, which makes the classification more effective and efficient. It works from the top node to the bottom by following a decision tree principle, and each node can be viewed as a binary classification. To improve the performance, the top node is usually set as the classification between the two classes with the biggest gap; for example, for a 3-class classification problem, the top node will be set as 1 vs. 3, with the assumption that 1 represents the highest and 3 represents the lowest.

2.5.4. Error Correcting Output Codes

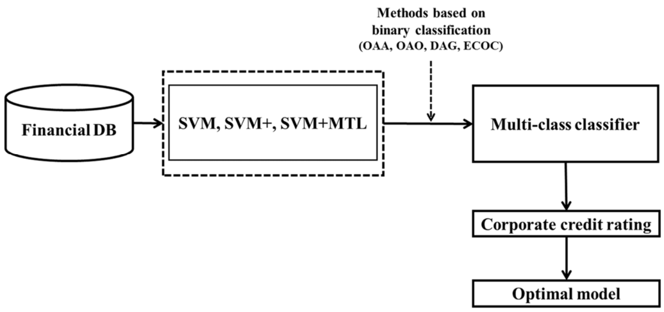
Dietterich and Bakiri (1995) used Error Correcting Output Codes (ECOC) to solve the multiclass learning problems. It works by constructing several classifiers: coding the objective class of a classifier as +1, and coding the others of a classifier as -1. The new instance will be labeled by calculating the hamming distance between the theoretical codes and the real predictive codes for each class, and the class with the minimum hamming distance will be selected as the predictive class. The prediction accuracy increases as the number of classifiers increases. However, the learning time will increase as the number of classifiers increases. Maximum classifiers are used for ECOC approach in the experiment to ensure the performance. Hence, as shown in <Table 2>, 7 classifiers are needed in total for a 4-class example. To some degree, the ECOC method can be considered as the extension of the OAA approach and hence it costs more computing time than the OAA approach.

Ⅲ. Research Framework

As the literature review introduced new data mining techniques, SVM+ and SVM+MTL (multi-task learning), they are competent alternative data mining techniques to corporate credit rating. Ribeiro et al. (2010), Ribeiro et al. (2012) firstly applied the new techniques to the bankruptcy prediction field and showed that SVM+ technique outperformed the conventional SVM. In addition, Ribeiro et al. (2010), Ribeiro et al. (2012) showed that SVM+ outperformed both SVM and SVM+MTL with an application to financial distress. This study applies the newly developed techniques SVM+ and SVM+MTL to the multi-classification and bond rating. The proposed model is the first application of SVM+ and SVM+MTL for multi-class classification and is designed to predict the corporate credit ratings with real cases in real world. That is, this paper aims to solve multi-class SVM+ and SVM+MTL problems in the context of bond rating. Also, we compare these new techniques with SVM, the most frequently-used method. In addition, this study intends to examine whether SVM+MTL outperforms all other techniques or not in each multi-class approach. The research framework is presented in <Figure 3>.

<Table 2> ECOC Method for a 4-Class Example (adapted from Kim and Ahn, 2012).

	OAA				(1,2) vs. (3,4)	(1,3) vs. (2,4)	(1,4) vs. (2,3)
Class 1	+1	-1	-1	-1	+1	+1	+1
Class 2	-1	+1	-1	-1	+1	-1	-1
Class 3	-1	-1	+1	-1	-1	+1	-1
Class 4	-1	-1	-1	+1	-1	-1	+1



<Figure 3> Research Framework

The concrete steps of the research are illustrated as follows:

(Step 1) To select the financial ratios and corporate ratings. For SVM+ and SVM+MTL, the group information variable also needs to be selected in this step.
(Step 2) To map the ready-processed data into SVM, SVM+, and SVM+MTL model.
(Step 3) To construct multi-class classifiers for SVM, SVM+ and SVM+MTL. Because SVM+ and SVM+MTL techniques are all originally designed for binary classification, methods that are based on binary classification and widely used to solve multi-class SVM problems are applied. To validate the robustness of SVM+ and SVM+MTL's performance, all the four basic multi-class techniques OAA, OAO, DAG, and ECOC are applied. Thereby, it can effectively prevent the contingency effect of the performance caused by a single approach and hence increase the explanatory power, stability, and persuasiveness of the proposed model.
(Step 4) To make corporate credit rating prediction. 12 methods (OAASVM, OAASVM+, OAASVM+ MTL, OAOSVM, OAOSVM+, OAOSVM+MTL, DAGSVM, DAGSVM+, DAGSVM+, DAGSVM+ MTL, ECOCSVM, ECOCSVM+, ECOCSVM+MTL) are applied in our experiment in total.
(Step 5) To select the optimal model. Optimization is realized by comparing all the algorithms for each approach and generating one optimal model for each approach and then selecting the best model across all the four basic multi-class approaches.

Ⅳ. Experiments and Analysis

4.1. Datasets

This study collects 5 years (2007-2011) financial data of 1716 companies listed in Korea stock market from FnGuide database, which is presented by Korea Information Service (KIS), a professional rating agency in Korea. Also, the corresponding 5 years' (2008-2012) rating data published by KIS is collected as the target predictive variable. After preprocessing the missing instances and outliers, we get 803 instances with 4 classes.

In this paper, we do not make a feature selection from a large amount (200 or 300) of variables, but use the significant variables used by the domain expert instead, because the combination of domain knowledge and data mining techniques make more accurate and persuasive prediction. Moreover, the collection of variables is not an easy work in most cases. Here, we follow Altman (2004) to select variables: WK/TA, RE/TA, EBIT/TA, ME/BL, Size and Age. Where WK is the net working capital, RE is retained earnings, TA is total assets, EBIT is earnings before interest and taxes, ME is the market value of equity, and BL is the book value of total liabilities. Size equals to total liabilities normalized by the total value of Korea Stock Exchange market. Then, we log-transform the variables as follows: RE/TA→-ln(1-RE/TA), EBIT/TA→-ln(1-EBIT/TA), ME/BL→1+ln(ME/BL), since log-transformation can reduce the skewness in the distribution of these variables.

For these ratings, to avoid the scarcity of each class, we set AAA and AA as class 1 (highly safe), A as class 2 (safe), BBB as a class 3 (less safe), BB, B, CCC and D as class 4 (risky). Mention to say, the premise or assumption behind the study is that there exists no credit default swap. That is, we believe that the data reliability. Regarding the group information, we select asset turnover ratio as the group information and split the instances into three groups with it.

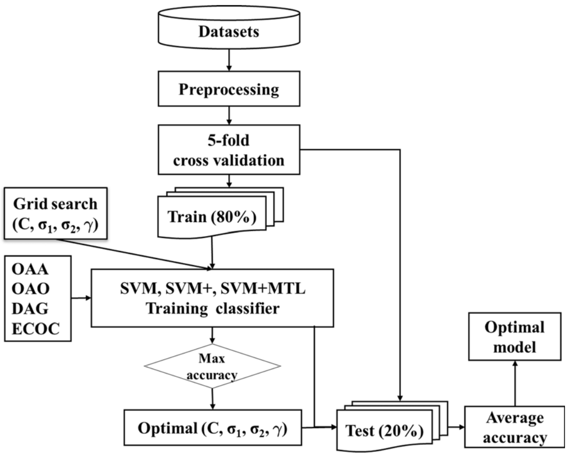
4.2. Experimental Design

The whole experimental process is presented in <Figure 4>. The whole dataset can be partitioned into training data set (80%) and test data set (20%) by utilizing five-fold cross validation. Improper selection of parameters can be moderated by using grid search. We applied the SVM, SVM+, and SVM+MTL techniques separately first, and then generated the optimal model by comparing the performance of all methods in each multi-class approach and simultaneously validate whether or not our proposed SVM+MTL model outperforms all other multi-class approaches (OAA, OAO, DAG, and ECOC). We ran SVM, SVM+, and SVM+MTL method with Matlab 2013a. The detailed experimental design for each approach is presented as follows.

For SVM, it is well known that RBF kernel function outperforms the other kernel functions (Huang et al., 2004; Kim, 2003; Lee, 2007; Tay and Cao, 2001). Hence, for SVM+ and SVM+MTL, we employ RBF kernel function both in decision space and correcting space.

Because there is no well-accepted structured way to select the parameters for SVM, a selection of parameters often can be done with experiments with SVM (Tay and Cao, 2001). For example, by showing over-fitting or under-fitting cases with the improper selection of parameters, Tay and Cao (2001) illustrated that kernel parameters should be carefully chosen due to its direct impact on the SVM performance. Hence, the parameter selection is an important issue in this study. To avoid the improper parameter selection, we use a grid search for the four parameters at the cost of time consuming to find an optimal set of parameters with the specified scope.

Specifically, a grid-search strategy is used in our study to explore the optimal parameters of C, σ_1, σ_2, γ (Note: C represents the penalty factor, σ_1 represents the RBF kernel function parameter in decision function, σ_2 represents the RBF kernel function parameter in correcting function, γ represents the weighting factor between decision function and correcting function in correcting function). For SVM,



<Figure 4> Experimental Process

a grid-search to explore the optimal parameters of C, σ_1; while for SVM+ and SVM+MTL, a grid-search to explore the optimal parameters of C, σ_1, σ_2, γ. Based on the study of Tay and Cao (2001), we set the parameters C and σ in the range (10, 100) and (1, 100), respectively. While the parameter γ is set in the range (0.001, 10) based on Liang (2009). So, the parameter C is set as (10, 25, 50, 75, 100), the parameter σ_1 and σ_2 are set as (1, 5, 7, 9, 10), the parameter γ is set as (10, 1, 0.1, 0.01, 0.001).

4.3. Results Analysis

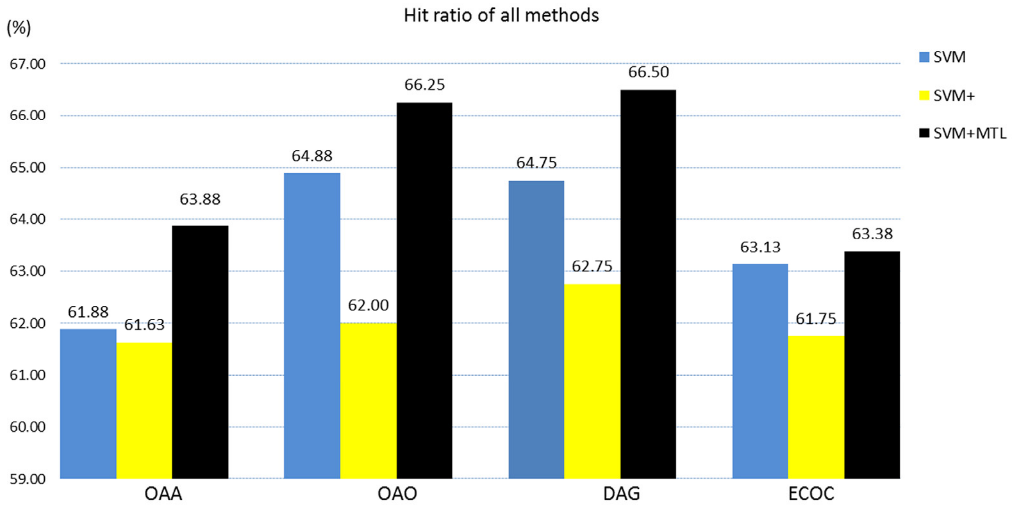
Here, we adopt the hit-ratio as the performance measure. The results of all methods are presented in <Table 3> and <Figure 5>. The results of SVM,

<Table 3> Experimental Results of All Methods

Algorithms	Partition	Set 1(%)	Set 2(%)	Set 3(%)	Set 4(%)	Set 5(%)	Average (%)
(1) OAA approach							
SVM	Train	75.27	75.74	78.85	75.12	76.05	76.21
	Test	60.00	59.38	62.50	64.38	63.13	61.88
SVM+	Train	73.41	81.96	67.81	64.07	80.09	73.47
	Test	65.00	61.25	61.25	52.50	68.13	61.63
SVM+MTL	Train	80.25	80.40	85.38	77.45	76.83	80.06
	Test	55.63	63.13	68.13	66.88	65.63	63.88
(2) OAO approach							
SVM	Train	73.41	80.40	62.52	72.16	73.87	72.47
	Test	63.75	63.75	64.38	63.75	68.75	64.88
SVM+	Train	77.14	74.34	73.87	63.92	67.50	71.35
	Test	60.00	61.25	65.00	65.63	58.13	62.00
SVM+MTL	Train	76.52	78.07	84.29	80.25	77.14	79.25
	Test	65.00	62.50	66.88	66.25	70.63	66.25
(3) DAG approach							
SVM	Train	74.49	80.09	63.45	72.32	72.63	72.60
	Test	63.75	64.38	64.38	63.13	68.13	64.75
SVM+	Train	77.92	73.56	74.18	65.79	67.50	71.79
	Test	59.38	63.13	65.63	64.38	61.25	62.75
SVM+MTL	Train	76.05	78.54	84.29	80.40	76.83	79.22
	Test	65.00	61.88	66.88	66.88	71.88	66.50
(4) ECOC approach							
SVM	Train	74.34	76.52	74.65	70.76	74.03	74.06
	Test	61.25	60.00	63.13	64.38	66.88	63.13
SVM+	Train	79.78	80.87	69.52	68.58	80.25	75.80
	Test	66.88	62.50	61.88	61.25	56.25	61.75
SVM+MTL	Train	74.96	79.00	84.76	73.41	77.92	78.00
	Test	56.88	61.25	66.88	66.25	65.63	63.38

SVM+, and SVM+MTL are explained in each multi-class approach. In OAA approach, the mean values of test accuracy for SVM and SVM+ are 61.88% and 61.63%, respectively. SVM+MTL algorithm outperforms all the other methods with 63.88% and yields the best performance. Yet SVM+ does not result in a better performance than conventional SVM. As shown in <Table 3>, the results of OAO approach show that SVM+MTL outperforms all the other methods with an accuracy of 66.25%, and does not show that SVM+ outperforms SVM either. Compared to OAA approach, OAO approach appears to be more effective. DAG approach takes the same training time as OAO approach, yet the amount of testing time is less than OAO approach and thus performs more effectively. As shown in <Table 2>, the average test results for SVM, SVM+, and SVM+MTL are 64.75%, 62.75% and 66.50% respectively. Similar to OAA and OAO approach, SVM+MTL performs best and SVM+ does not perform better than SVM. The results also show that the DAG approach outperforms the other approaches for the corresponding algorithm (i.e., when comparing each algorithm respectively). This may benefit from its unique effective model design. ECOC is the most time consuming approach of all the multi-class processing approaches. Likewise, SVM+MTL method obtains the best performance. ECOC is not a suggested method for its too long computing time. Theoretically, ECOC needs 3 times of amount of training time than OAA approach.

Taken together, we choose SVM+MTL method as the optimal algorithm for the four multi-class approaches by comparing hit-ratio. Here, we also list the experimental time for a reference. The average amount of computing time for each multi-class approach is listed in the ascending order, from the shortest to the longest, DAG < OAO < OAA < ECOC. And for each approach, there exists the relationship in the average computing time of each technique: SVM < SVM+MTL < SVM+. Compared to SVM+ and SVM+MTL techniques, SVM needs much less time for the training time. The computing time of SVM+MTL is less than SVM+ across all 4 approaches.



<Figure 5> Hit Ratio of All Methods

In addition, we list the within-1-class accuracy results of each approach, which presents the probabilities of the predictions to be within one class away from the actual rating, so as to compare the prediction accuracy by allowing a certain error. As shown in <Table 4>, the within-1-class prediction probabilities are all over 90%. SVM+MTL method has the best within-1-class prediction accuracy among them.

<Table 4> Within-1-Class Accuracy of All Methods

Actual Rating	Predicted Rating											
	OAASVM				OAASVM+				OAASVM+MTL			
	1	2	3	4	1	2	3	4	1	2	3	4
1	164	62	3	1	179	45	5	1	177	48	4	1
2	57	171	23	14	74	159	23	9	66	156	31	12
3	5	68	28	39	20	57	43	20	2	56	48	34
4	2	18	13	132	11	21	21	112	1	15	19	130
Accuracy	94.63%				91.63%				95.63%			

Actual Rating	OAOSVM				OAOSVM+				OAOSVM+ MTL			
	1	2	3	4	1	2	3	4	1	2	3	4
1	172	54	4	0	167	60	2	1	184	41	5	0
2	62	177	18	8	59	154	45	7	59	164	35	7
3	6	64	45	25	8	53	56	23	3	55	60	22
4	3	15	22	125	1	18	27	119	0	15	28	122
Accuracy	95.50%				95.38%				96.25%			

Actual Rating	DAGSVM				DAGSVM+				DAGSVM+ MTL			
	1	2	3	4	1	2	3	4	1	2	3	4
1	167	59	4	0	159	66	4	1	180	44	6	0
2	57	177	23	8	46	160	52	7	52	167	39	7
3	3	63	49	25	4	49	64	23	2	53	63	22
4	1	15	24	125	1	12	33	119	0	11	32	122
Accuracy	96.13%				96.38%				96.75%			

Actual Rating	ECOCSVM				ECOCSVM+				ECOCSVM+MTL			
	1	2	3	4	1	2	3	4	1	2	3	4
1	174	53	2	1	183	43	2	2	189	39	1	1
2	65	182	9	9	78	154	25	8	78	167	15	5
3	7	88	25	20	17	64	41	18	7	76	33	24
4	4	26	11	124	4	31	14	116	3	27	17	118
Accuracy	93.88%				92.00%				94.50%			

Ⅴ. Conclusion

In this study, we applied SVM+ and SVM+MTL techniques to multi-classification and to corporate credit rating prediction. By comparing the experiment results of all techniques, the objective is to select the optimal model. To avoid the contingency of the performance caused by a single approach, we applied the four kinds of multi-class approaches OAA, OAO, DAG, and ECOC to SVM, SVM+ and SVM+MTL techniques, separately. The results show the superiority of SVM+MTL to all the other techniques in each multi-class approach. Finally, we chose SVM+MTL technique as the optimal algorithm by comparing hit-ratio results. The additional benefit of our credit rating approach is to show more detailed explanatory powers by transforming a binary bankruptcy prediction problem into multi-class credit rating analysis.

Additionally, our results also show that DAG is the most effective and efficient approach in the four multi-class approaches. Both ECOC and OAA approach are time consuming undoubtedly, especially when the number of samples is very large. ECOC approach needs much more computing time than OAA approach. SVM+ and SVM+MTL techniques are more time consuming than SVM, since more tuning parameters are needed. Hence, the computing speed for SVM+ and SVM+MTL still needs to be improved.

From our experimental results, it is observed that the group information (asset turnover ratio) selected in this study is appropriate for our dataset. The group information may have a critical impact on the result of SVM+ and SVM+MTL techniques. As referred in Liang et al. (2009), the result may perform poor with an improper selection of group information. However, it does not seem too difficult and can be avoided with deep domain knowledge.

Our study validates the outstanding performance of newly developed techniques SVM+ and SVM+MTL in Korea bond rating. Most importantly, this study proposes an alternative solution for multi-classification and corporate bond rating prediction as we show SVM+MTL technique improves the prediction accuracy significantly.

As all research, this study is not exceptional for limitations, which can also suggest future research avenues and extensions. In this study, single group information is selected regarding computing time, so it may ignore the impact of varying group information on the results. Moreover, future study can modify the range of each parameter to find a better grid search. To validate whether the group information is not an improper selection, a work can be done by applying several different group information.

<References>

[1] Altman, E. I., and Katz, S. (1976). Statistical Bond Rating Classification Using Financial and Accounting Data. *In Proceedings of the Conference on Topical Research in Accounting.* New York University School of Business New York, 205-239.

[2] Altman, E. I., and Rijken, H. A. (2004). How Rating Agencies Achieve Rating Stability. *Journal of Banking & Finance, 28*(11), 2679-2714.

[3] Bellotti, T., and Crook, J. (2009). Support Vector Machines for Credit Scoring and Discovery of Significant Features. *Expert Systems with Applications, 36*(2), 3302-3308.

[4] Bennell, J. A., Crabbe, D., Thomas, S., and Ap Gwilym, O. (2006). Modelling Sovereign Credit Ratings:

Neural Networks Versus Ordered Probit. *Expert Systems with Applications, 30*(3), 415-425.

[5] Cao, L., Lim, K. G., and Zhang, J. (2006). Bond Rating Using Support Vector Machine. *Intelligent Data Analysis, 10*(3), 285-296.

[6] Chen, W. H., and Shih, J. Y. (2006). A Study of Taiwan's Issuer Credit Rating Systems Using Support Vector Machines. *Expert Systems with Applications, 30*(3), 427-435.

[7] Crammer, K., and Singer, Y. (2000). On the Learn Ability and Design of Output Codes for Multiclass Problems. *In Proceedings of the 13th Annual Conference on Computational Learning Theory,* Palo Alto, California, 35-46.

[8] Dietterich, T., and Bakiri, G. (1995). Solving Multiclass Learning Problems Via Error-Correcting Output Codes. *Journal of Artificial Intelligence Research, 2,* 263-286.

[9] Dutta, S., and Shekhar, S. (1988). Bond Rating: A Non-Conservative Application of Neural Networks. *In Proceedings of IEEE International Conference on Neural Networks,* San Diego, California, 443-450.

[10] Fisher, L. (1959). Determinants of Risk Premiums on Corporate Bonds. *Journal of Political Economy, 67*(3), 217-237.

[11] Hong, T., and, Park, J. (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. *Asia Pacific Journal of Information Systems, 21*(2), 43-58.

[12] Horrigan, J. O. (1966). The Determination of Long-Term Credit Standing with Financial Ratios. *Journal of Accounting Research, 4,* 44-62.

[13] Hsu, C. W., and Lin, C. J. (2002). A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks, 13*(2), 415-425.

[14] Hsu, C. F., and Hung, H. F. (2009). Classification Methods of Credit Rating-A Comparative Analysis on SVM, MDA and RST. *In Computational Intelligence and Software Engineering, CiSE 2009, International Conference, IEEE,* 1-4.

[15] Huang, C. L., Chen, M. C., and Wang, C. J. (2007). Credit Scoring with a Data Mining Approach Based on Support Vector Machines. *Expert Systems with Applications, 33*(4), 847-856.

[16] Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., and Wu, S. (2004). Credit Rating Analysis with Support Vector Machine and Neural Networks: A Market Comparative Study. *Decision Support Systems, 37*(4), 543-558.

[17] Kamstra, M., Kennedy, P., and Suan, T. K. (2001). Combining Bond Rating Forecasts Using Logit. *Financial Review, 36*(2), 75-96.

[18] Kim, J. W. (1993). Expert Systems for Bond Rating: A Comparative Analysis of Statistical Rule-Based and Neural Network Systems. *Expert Systems, 10*(3), 167-172.

[19] Kim, K. S., and Han, I. (2001). The Cluster-Indexing Method for Case-Based Reasoning Using Self-Organizing Maps and Learning Vector Quantization for Bond Rating Cases. *Expert Systems with Applications, 21*(3), 147-156.

[20] Kim, K. J., and Ahn, H. C. (2012). A Corporate Credit Rating Model Using Multi-Class Support Vector Machines with an Ordinal Pairwise Partitioning Approach. *Computers & Operations Research, 39*(8), 1800-1811.

[21] Kumar, K., and Bhattacharya, S. (2006). Artificial Neural Network vs Linear Discriminant Analysis in Credit Ratings Forecast: A Comparative Study of Prediction Performances. *Review of Accounting and Finance, 5*(3), 216-227.

[22] Kwon, Y. S., Han, I. G., and Lee, K. C. (1997). Ordinal Pairwise Partitioning (OPP) Approach to Neural Networks Training in Bond Rating. *Intelligent Systems in Accounting, Finance and Management, 6,* 23-40.

[23] Lee, T. S., Chiu, C. C., Lu, C. J., and Chen, I. F. (2002). Credit Scoring Using the Hybrid Neural Discriminant Technique. *Expert Systems with Applications, 23*(3), 245-254.

[24] Lee, Y. C. (2007). Application of Support Vector Machines to Corporate Credit Rating Prediction.

Expert Systems with Applications, 33(1), 67-74.
[25] Liang, L., Cai, F., and Cherkassky, V. (2009). Predictive Learning with Structured (Grouped) Data. *Neural Networks, 22*(5), 766-773.
[26] Maher, J. J., and Sen, T. K. (1997). Predicting Bond Ratings Using Neural Networks: A Comparison with Logistic Regression. *Intelligent Systems in Accounting, Finance and Management, 6*(1), 59-72.
[27] Park, J., and Hong, T. (2009). The Prediction of DEA based Efficiency Rating for Venture Business Using Multi-class SVM. *Asia Pacific Journal of Information Systems, 19*(2), 139-155.
[28] Pinches, G. E., and Mingo, K. A. (1973). A Multivariate Analysis of Industrial Bond Ratings. *Journal of Finance, 28*(1), 1-18.
[29] Pinches, G. E., and Mingo, K. A. (1975). The Role of Subordination in Determining Industrial Bond Ratings. *Journal of Finance*, *30*(1), 201-206.
[30] Ribeiro, B., Silva, C., Chen, N., Vieira, A., and Neves, J. C. D. (2012). Enhanced Default Risk Models with SVM+. *Expert Systems with Applications, 39*(11), 10140-10152.
[31] Ribeiro, B., Silva, C., Vieira, A., Cunha, A., and Carvalho das Neves, J. (2010). Financial Distress Model Prediction Using SVM+. *In IEEE International Joint Conference on Neural Networks,* 1-7.
[32] Shin, K. S., and Han, I. G. (2001). A Case-Based Approach Using Inductive Indexing for Corporate Bond Rating. *Decision Support Systems, 32*(1), 41-52.
[33] Tay, F. E. H., and Cao, L. J. (2001). Application of Support Vector Machines in Financial Time Series Forecasting. *Omega*, *29*(4), 309-317.
[34] Thulasiraman, K., and Swamy, M. N. S. (1992). Acyclic Directed Graphs. *In J. K. Thulasiraman, M. N. S. Swamy* (Eds.), *Graphs: Theory and Algorithms.* John Wiley and Sons Inc, 118.
[35] Vapnik, V. N., and Vashist, A. (2009). A New Learning Paradigm: Learning Using Privileged Information. *Neural networks, 22*(5-6), 544-557.
[36] Vapnik, V. N. (1995). *The Nature of Statistical Learning theory.* New York, Springer-Verlag.
[37] Vapnik, V. N. (2006). *Estimation of Dependencies based on Empirical Data, In Empirical Inference Science (Second Edition).* New York, Springer.
[38] West, R. R. (1970). An Alternative Approach to Predicting Corporate Bond Ratings. *Journal of Accounting Research*, *8*(1), 118-125.
[39] Weston, J., and Watkins, C. (1999). Support Vector Machines for Multiclass Pattern Recognition. *In Proceedings of the Seventh European Symposium on Artificial Neural Networks.* Bruges, Belgium, 219-224.
[40] Yu, L., Wang, S., and Lai, K. K. (2008). Credit Risk Assessment with a Multistage Neural Network Ensemble Learning Approach. *Expert Systems with Applications, 34*(2), 1434-1444.

◆ About the Authors ◆

Gang Ren

Gang Ren is a Ph.D. Candidate of Management Information Systems at College of Business Administration, Pusan National University in Korea. His current research interests are data mining for financial application, opinion mining, and CRM.

Taeho Hong

Taeho Hong is a professor of Management Information Systems at College of Business Administration, Pusan National University in Korea. He received the Ph.D from Korea Advanced Institute of Science and Technology. He worked for Deloitte Consulting as a senior consultant. His research interest includes intelligent systems, data mining, and recommender systems for e-business. He has published his research in Expert Systems with application, Expert systems, Asian Journal of Information Systems.

YoungKi Park

YoungKi Park is an assistant professor of Information Systems at the George Washington University, School of Business. He received his PhD in business administration from the University of Southern California and MS from KAIST in Korea. He also has 10 years of IT consulting experience in business intelligence, analytics and enterprise systems at Oracle, LG, and the Korea Exchange. His research focuses on IS/IT Strategy in Turbulent Environments and he specializes in the set-theoretic configurational approach, fuzzy-set qualitative comparative analysis (fsQCA). He was selected for the ICIS 2009 Doctoral Consortium, and his work has been presented at several international conferences such as ICIS, HICSS, AOM, AMCIS, and DESRIST. His work has been published in Information Systems Research, Research in Sociology of Organizations, and InformationandSoftwareTechnology.

Submitted: December 31, 2014; 1st Revision: June 12, 2015; Accepted: July 10, 2015