Asia Pacific Journal of Information Systems
Vol. 20, No. 2, June 2010

Customer Level Classification Model Using Ordinal Multiclass Support Vector Machines*

Kyoung-jae Kim**, Hyunchul Ahn***

Conventional Support Vector Machines (SVMs) have been utilized as classifiers for binary classification problems. However, certain real world problems, including corporate bond rating, cannot be addressed by binary classifiers because these are multi-class problems. For this reason, numerous studies have attempted to transform the original SVM into a multiclass classifier. These studies, however, have only considered nominal classification problems. Thus, these approaches have been limited by the existence of multiclass classification problems where classes are not nominal but ordinal in real world, such as corporate bond rating and multiclass customer classification. In this study, we adopt a novel multiclass SVM which can address ordinal classification problems using ordinal pairwise partitioning (OPP). The proposed model in our study may use fewer classifiers, but it classifies more accurately because it considers the characteristics of the order of the classes. Although it can be applied to all kinds of ordinal multiclass classification problems, most prior studies have applied it to finance area like bond rating. Thus, this study applies it to a real world customer level classification case for implementing customer relationship management. The result shows that the ordinal multiclass SVM model may also be effective for customer level classification.

Keywords : Support Vector Machines, Ordinal Pairwise Partitioning, Multiclass Classification, Customer Level Classification, Customer Relationship Management

---

* This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2007-327-B00211).
** Main author, Department of Management Information Systems, Dongguk University-Seoul
*** Corresponding author, School of Management Information Systems, Kookmin University

Ⅰ. Introduction

Recently, customer relationship management (CRM) has been one of the most popular subjects in marketing research. Among many research themes vis-a-vis CRM, customer classification is considered one of the important issues. Customer classification means that a company classifies its customers into predefined groups with similar purchasing behavior patterns. In general, many companies construct customer classification models to determine product feasibility. This kind of knowledge may create many crucial marketing opportunities, such as one-to-one marketing, direct mailing, and sales promotion using telephone or e-mail. For this reason, many companies such as Ford, Allstate, and 1-800-flowers.com analyze their customers' profiles and purchasing behaviors and construct customer classification models based on the probability of customer purchase.

Traditionally, many statistical classification techniques, including multivariate discriminant analysis, Probit, and logistic regression, have used for the construction of customer classification models. Recently, support vector machines (SVMs) have become popular as a solution for classification problems because of their robustness and high classification performance. SVMs, however, were originally developed for binary classification problems. As such, they may not solve accurately multiclass classification problems, such as bond rating and multiclass customer classification. Many researchers have tried to transform the binary SVM into a multiclass classifier. There have been several studies in this regard. One such study constructs and combines several binary classifiers. Another directly considers all data in one optimization formulation. However, these approaches had only focused on classifying samples into nominal categories [Hsu and Lin, 2002a; Statnikov *et al*., 2005].

In this study, we use a novel multiclass SVM approach that can efficiently and effectively consider multiple classes with ordinal characteristics, such as customer's profitability levels. We employ the ordinal pairwise partitioning (OPP) approach for the proposed model. The OPP approach partitions a data set into sub data sets with reduced classes in an ordinal and pairwise manner according to output classes [Kwon *et al*., 1996]. The proposed model utilizes additional hidden information, the order of the classes, for classification. Thus, it is possible to get more accurate prediction results with fewer classifiers than with conventional multiclass SVMs [Ahn and Kim, 2006].

This technique can be applied to any kinds of ordinal multiclass classification problems. Nonetheless, the prior studies that adopted it as a tool for ordinal multiple classification are very rare. Especially, most of the prior studies that used ordinal multiclass classification techniques just have applied them to finance domains like predicting bond rating or credit rating. Thus, in this study, we apply it to a real world classification case for customers' profitability level for the CRM purpose. In addition, we compare the results of the model to those of multiple discriminant analysis (MDA), case-based reasoning (CBR), artificial neural networks (ANN), and other multiclass SVM algorithms.

The rest of this study is organized as follows. The next section reviews prior literatures on conventional binary SVMs and multiclass SVMs. In section 3, our approach for multiclass SVMs is explained. Section 4 describes the research data and the methodologies for the validation of our model. In Section 5, the experimental results are summarized and discussed. The final section presents the conclusions and future research direction of the study.

Ⅱ. Prior Studies

2.1 Prior Studies on Customer Classification and Response Modeling

During the past decade, many studies on customer classification have been performed. Customer classification is interchangeably referred to as response modeling, which predicts the probability of customer response to marketing promotion. Although, there is a slight difference between customer classification and response modeling, we consider prior studies on response modeling in our study. Traditionally, statistical techniques have been employed to perform response modeling. Malthouse [1999] implemented ridge regression and stepwise regression to construct response models. He showed that the former is more stable than the latter. Colombo and Jiang [1999] suggested the Recency-Frequency-Monetary (RFM) Analysis as a tool for response modeling.

Recent research tends to use data mining techniques such as decision tree, case-based reasoning (CBR), and artificial neural networks (ANNs) in this research area. ANNs have popularly been used in response modeling [Moutinho *et al*., 1994; Bounds and Ross, 1997; Potharst *et al*., 2001]. In addition, Ha *et al*. [2005] proposed bagging neural network model, a variant of ANN, as a tool for response modeling. However, some studies reported that ANNs do not outperform statistical techniques [Suh *et al*., 1999; Zahavi and Levin, 2000]. Decision tree has also been one of the most popular techniques for response modeling. Haughton and Oulabi [1997] applied classification and regression tree (CART) and CHAID to response modeling. Ling and Li [1998] also proposed the Naïve Bayes model and C4.5 as response modeling techniques.

Some researchers applied CBR to response modeling. Coenen *et al*. [2000] employed a combined rule-induction and case-based reasoning model to predict prospects. They initially classified buyers and non-buyers through a C5 algorithm and then integrated it with case-based reasoning to generate final results. Chiu [2002] used hybrid CBR and the genetic algorithms (GAs) model to predict customers' response. He used GA to search the best set of weighting values in CBR. He reported that his model could outperform the regression-based CBR in classification performance. In addition, Ahn *et al*. [2006] proposed the simultaneous optimization model for CBR using GA to classify customer profitability level. They used GA as an optimization tool for instance selection and feature weighting in CBR and compared it with Logit, MDA, ANN, and binary SVM. They proposed that the simultaneous model could perform better than other comparative models in addressing customer classification problems. Ahn *et al*. [2007] also employed GA to reduce two dimensions of customer database for customer classification models.

More recently, SVM has been applied to a wide variety of application domains, including customer classification and response modeling. Viaene *et al*. [2001] proposed the least squares SVM (LS-SVM) to predict purchase incidence. They extended beyond the standard RFM modeling semantics by including alternative operationalizations of the RFM variables and by adding several non-RFM variables. They reported that the elimination of redundant or irrelevant input variables allowed for a significant reduction in model complexity. Shin and Cho [2006] presented a way to alleviate problems with large training data using a novel informative sampling and provided guidelines for limitations with class imbalance and scoring from binary SVM output for response modeling. Lee and Cho [2007] proposed to use novelty detection approaches to alleviate class imbalance problems in response modeling. They used one-class SVM and learning vector quantization for novelty detection and compared them with binary classifiers for catalogue mailing data. They showed that SVM with modified misclassification costs perform best when response rates are relatively high. In addition, Kim *et al*. [2008] employed support vector regression to estimate the purchase amounts of respondents. They also used ε-tube based sampling to support vector regression that led to better accuracy than a random sampling method did.

Although above mentioned prior studies proposed various methods to predict prospects, most of them did not consider the problems associated with multiclass classification in customer profitability level. In addition, there were many studies on multiclass classification for corporate bond rating, but few for customer classification problems. Thus, there has been a growing research interest vis-a-vis multiclass customer classification problems.

2.2 Conventional Support Vector Machines

Conventional binary SVM uses a linear model to implement nonlinear class boundaries through certain nonlinear transformation of input space into high-dimensional feature space. A linear model constructed in a new high-dimensional feature space can represent a nonlinear decision boundary from the original space. An optimal separating hyperplane is constructed in a new feature space [Vapnik, 1995].

For this reason, SVM is considered an algorithm that finds a special kind of linear model, the *maximum margin hyperplane*. It allows the maximum separation between decision classes. The training data that are closest to the maximum margin hyperplane are called *support vectors*.

As mentioned above, SVM constructs a linear model to implement nonlinear class boundaries through the transformation of inputs into high-dimensional feature space. The kernel function does this work. There are different kernels for generating inner products to build machines with different types of nonlinear decision surfaces in the input feature space. Some examples of such kernel functions include the following: the linear kernel, $K(x, y) = x^T y$; the polynomial kernel, $K(x, y) = (x^T y + 1)^d$, where d is the degree of the polynomial kernel; and the Gaussian radial basis function (RBF), $K(x, y) = \exp(-1/\sigma^2 (x-y)^2)$, where σ^2 is the bandwidth of the function. The best model can be selected by choosing among different kernels the model that minimizes the estimate [Tay and Cao, 2001].

2.3 Multiclass Support Vector Machines

As mentioned above, SVMs were originally developed for binary classification problems. Thus, there has been a need to effectively transform SVMs so as to enable them to accommodate multiclass classification problems. Many prior studies have proposed several variants of the original SVM for implementing multiclass classification problems. To-date, there have been two types of approaches for extending SVMs to multiclass problems. One has been to decompose multiclass problems into several binary sub-problems, then solving each problem at once and compositing them. That is, multiclass SVMs can be implemented by building and combining several binary SVM classifiers. Second has been to consider all data at once in a single optimization formulation. A modification of the typical algorithm of SVM is required in this case. There are different techniques for each approach, and <Table 1> presents an overview of these techniques [Lorena and de Carvalho, 2008; Wu *et al*., 2008].

The first approach, *Constructing Several Binary Classifiers and Combining Them*, is divided into two sub-approaches. The first sub-approach, *One-Against-All*, is the simplest multiclass classification method. It first constructs k conventional binary SVM classifiers for k-class classification, such as "class 1 versus all other classes," "class 2 versus all other classes," and "class k versus all other classes" [Kreβel, 1999]. Then, it combines the results of k classifiers. For the case of "class 1 versus all other classes," the combined One-Against-All decision function selects the class of a sample that corresponds to the maximum value of k binary classification functions specified by the furthest "class 1" hyperplane. The hyperplanes calculated by k SVMs shift by doing so [Statnikov *et al*., 2005].

The second sub-approach, One-Against-One, constructs binary SVM classifiers for all pairs of classes. Finally, there are ${}_kC_2 = \frac{k(k-1)}{2}$ pairs. This means that for each pair of classes, a binary SVM classifier is constructed by solving the underlying optimization problem for maximizing the margin between the two classes in the pair. This approach follows the so-called *Max Wins* strategy. Thus, the decision function as-

<Table 1> Multiclass SVM Techniques Proposed in Literature

Approach	Proposed Technique	No. of classifiers to be constructed (Assumed k classes problem)	Reference
Constructing Several Binary Classifiers and Combining Them	One-Against-All	k	Kreβel [1999]
	One-Against-One	${}_kC_2$	Friedman [1996]
Considering All the Data at Once	Weston and Watkins	1	Weston and Watkins [1999]

signs a sample to a class that has the largest number of votes. When a tie occurs, a sample will be assigned a label that is based on the classification that is yielded by the furthest hyperplane [Friedman, 1996; Kreβel, 1999; Statnikov *et al*., 2005].

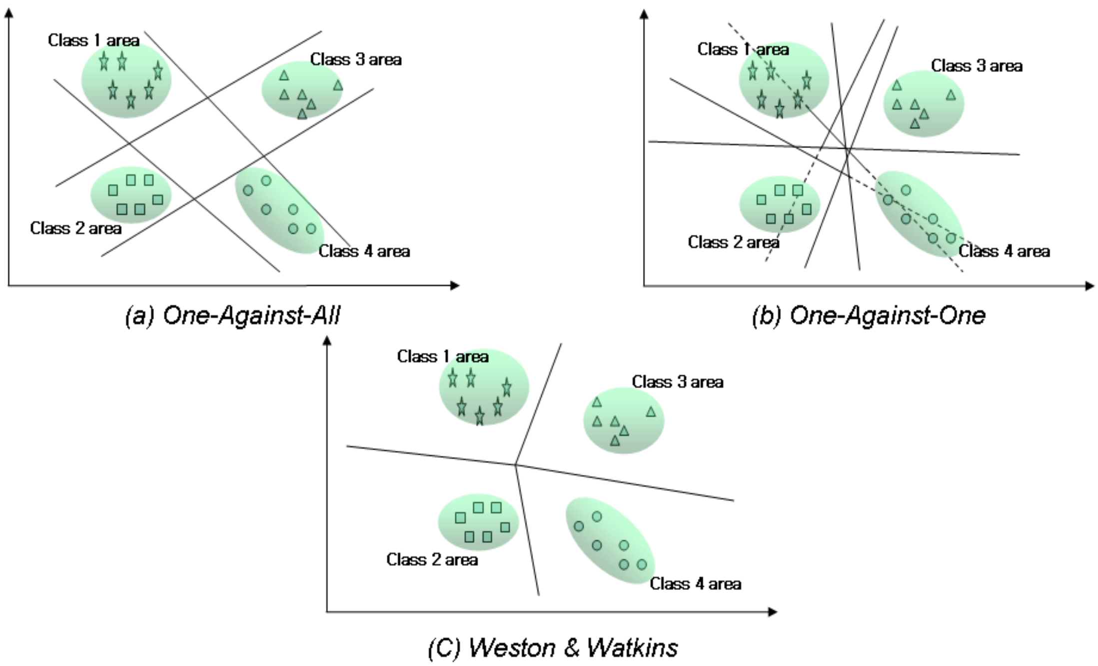
The second approach is *Considering All Data At Once*. There are several methods proposed by researchers; we use the method by *Weston and Watkins* in this study [Weston and Watkins, 1999]. This approach can be recognized as a simple extension of the binary SVM classification problem. If there is a k-class case, we solve single quadratic optimization problem of size *(k-l)n,* which is identical to a binary SVM for the case of k=2 in this approach [Hsu and Lin, 2002a; Hsu and Lin, 2002b; Platt, 1999].

<Figure 1> shows the differences among these three approaches in a four-class classification problem.

2.4 Ordinal Multiclass SVMs

As indicated in the previous section, there have been several approaches to transform binary SVMs into multiclass SVMs. All of these methods, however, have had a common limitation: all of them were developed for the multiclass classification problems where decision classes are nominal. Therefore, they may not appropriately address certain multiclass classification problems where decision classes are ordinal (i.e., decision classes have orders).

In this study, we employ an ordinal multiclass SVM model that uses the ordinal pairwise partitioning (OPP) approach as a tool for extending the conventional multiclass SVM models in order to address ordinal classes properly.



<Figure 1> Graphic Example of Three Approaches for Multiclass Support Vector Machines

The OPP approach partitions a data set into groups of sub-data sets with reduced classes in an ordinal and pairwise manner in accordance with output classes [Kwon *et al.*, 1996]. The ordinal multiclass SVM model was first proposed by Ahn and Kim[2006], and we will apply the model to our customer classification problem.

The processes of the ordinal multiclass SVMs are as follows: there are two partitioning and chaining methods. The partitioning methods consist of the One-Against-The Next and One-Against-Followers methods. The first partitioning method, One-Against-The-Next, is the (N-1) vs. N style, where, for example, N is 2, 3, 4 in the case of 4-class classification problem. It is similar to One-Against-One in conventional multi-class SVMs, but more efficient.

The One-Against-The-Next method only needs k-1 binary classifiers, where k is the total number of classes because it considers the order of decision classes. On the other hand, the One-Against-One method requires binary classifiers for each pair of classes. For example, for a classification problem with four decision classes, the One-Against-One method needs six binary classifiers (i.e. pairs of (1 class vs. 2 class), (1 vs. 3), (1 vs. 4), (2 vs. 3), (2 vs. 4), (3 vs. 4)). On the other hand, the One-Against- The Next method only requires three binary classifiers: (1 class vs. 2 class), (2 vs. 3), and (3 vs. 4).

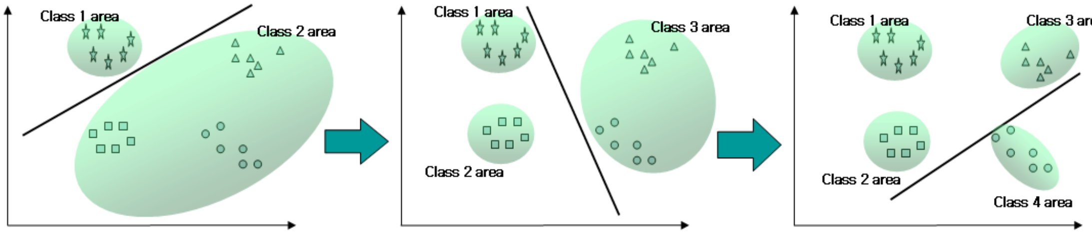
The second partitioning method is the (N) vs. (remaining classes) style, which we call the "One-Against-Followers approach." This is similar to One-Against-All of conventional multiclass SVMs, but more efficient than One-Against-All. Although One-Against-All constructs k binary classifiers, One-Against-Followers builds only k-1 binary classifiers when there are k decision classes. For example, for a classification problem with four decision classes, the One-Against-All method needs four binary classifiers, i.e., pairs of (1 class vs. 2&3&4 class), (2 vs. 1&3&4), (3 vs. 1&2&4), and (4 vs. 1&2&3). However, the One-Against-Followers method only requires three binary classifiers: (1 class vs. 2&3&4 class), (2 vs. 3&4), and (3 vs. 4).

As we can see from the above, any kind of ordinal SVMs requires just k-1 binary classifiers to classify data into k classes, while other conventional SVMs require from k to $k(k\text{-}1)/2$ models.

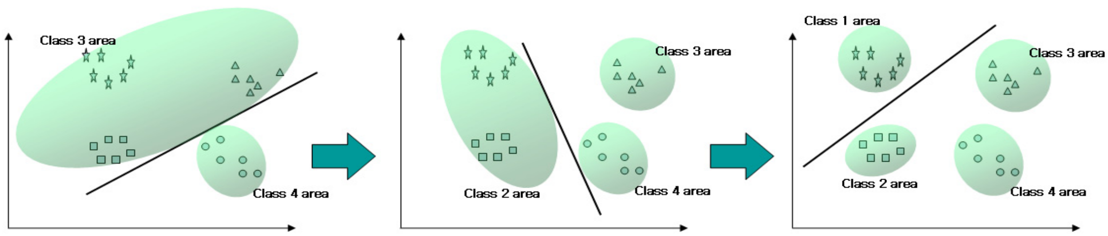
In addition, there are forward and backward methods regarding methods of fusion. The former determines the highest level of classes first and the lowest level classes last, and the latter is the opposite.

Finally, we can combine two partitioning methods and two fusing methods and then obtain four combined methods: "One-Against-The-Next + Forward method," "One-Against-The-Next + Backward method," "One-Against-Followers + Forward method," and "One-Against-Followers + Backward method."

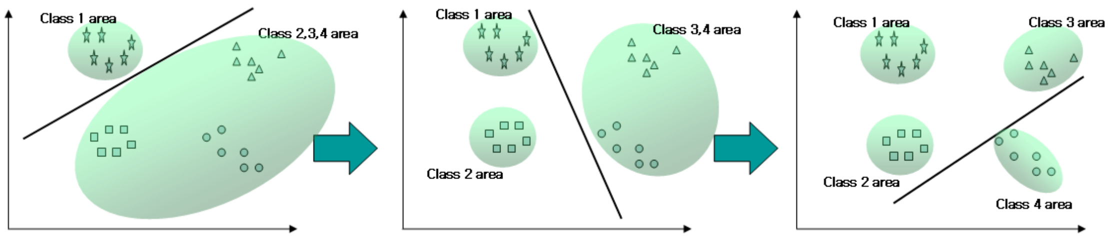
In our customer classification problem, we have four profitability classes (i.e., 1～4 class). For the One-Against-The-Next approach, using forward and backward computation, the data set is partitioned in advance. When using the forward method, we pair them to make three separate data sets: (1 vs. 2), (2 vs. 3), and (3 vs. 4). This is "One-Against-The-Next + Forward approach." When using the backward method, we pair them three data sets: (4 vs. 3), (3 vs. 2), and (2 vs. 1). This is "One-Against-The-Next + Backward approach." In the One-Against-Followers approach, when using the forward method, the data set is partitioned into pairs, such as pair of (1 vs. 2&3&4), a pair of (2 vs. 3&4), and a pair of (3 vs. 4). This is "One-Against-Followers + Forward approach." When using the backward method, the data set is partitioned to make a pair of (4 vs. 3&2&1), a pair of (3 vs. 2&1), and a pair of (2 vs. 1). This is "One-Against-Followers + Backward approach." <Figure 2> shows the differences among the ordinal multiclass SVM models.



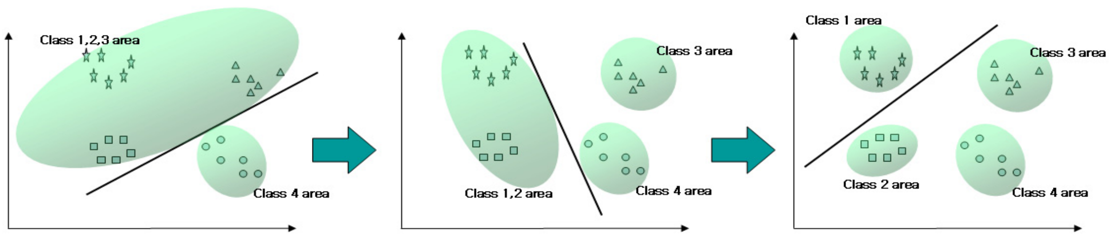
(a) One-Against-The-Next + Forward approach



(b) One-Against-The-Next + Backward approach



(c) One-Against-Followers + Forward approach



(d) One-Against-Followers + Backward approach

<Figure 2> Differences Among Four Types of the Ordinal Multiclass SVMs: Graphical Presentation

Ⅲ. Experimental Design and Results

3.1 Research Data

In order to validate our model, we applied the proposed model to a real world customer level classification data from a telecommunication service provider in Korea. Recently, the Korean government's policy of Mobile Number Portability (MNP) in the mobile communication industry has provoked severe competition among the service providers. As a result, many customers have changed their mobile service providers, and the mobile service providers have much interest in finding valuable customers among these incoming customers. Thus, it becomes very important for mobile service providers to build an effective customer classification model that can predict an incoming customers' expected value (i.e. his or her expected customer level) by using several demographic data and the short-term usage data. Our target company also has interests in building the customer classification model that classifies its existing customers or prospects into several levels according to their profitability.

The research data included 4,000 cases that consisted of various input variables on customers' demographic information and call details. Data also consisted of 38 independent variables and a dependent variable that represents the level of customer profitability. The company has classified their customers into 4 levels according to their profitability: VIP, gold, silver, and bronze. We used this as the dependent variable for our study. Among the initial variables, we chose 31 independent variables that affected the dependent variable with statistical significance at the 95% level by applying the independent samples t-test and chi-square test. Finally, we chose only 14 independent variables that had proven to be most influential to the customer profitability level. We did this by using the stepwise selection procedure with Wilk's λ in multivariate discriminant

<Table 2> Selected Independent Variables and Their Descriptions

Variable name	Description
BP1	Type of billing program is "A"
BP2	Type of billing program is "B"
BP3	Type of billing program is "C"
BP4	Type of billing program is "D"
HS1	The maker of the user's cell phone is "X" company.
HS4	The maker of the user's cell phone is "Y" company.
HS6	The maker of the user's cell phone is "Z" company.
BillProgValid	Total charge is higher than basic charge.
CallLenDay	The length of total calls during the day time
CallLenIntern	The length of international calls
AvgCallLenDom	The average length of domestic calls
TotCallLen	Total length of calls
PercentDay	The portion of domestic calls during the day time
PercentIntern	The portion of international calls relative to domestic calls

<Table 3> The Number of Each Data Set

Data set	Profitability level of the customers	The number of data	Proportions
Train	VIP Gold Silver Bronze	600 600 600 600	15% 15% 15% 15%
Test	VIP Gold Silver Bronze	200 200 200 200	5% 5% 5% 5%
Hold-out	VIP Gold Silver Bronze	200 200 200 200	5% 5% 5% 5%
Total		4,000	100%

analysis. The F value for stepwise entry was set at 3.84 and that for stepwise removal was set at 2.71. <Table 2> describes the selected independent variables.

To test the generalization of the proposed model, we divided the data into the three groups: training, test, and hold-out data sets. <Table 3> shows the number and proportion of samples in each data set.

3.2 Experimental Design

In this section, we examine our ordinal multiclass SVM model and compare it with multivariate discriminant analysis (MDA), case-based reasoning (CBR), artificial neural network (ANN), and other conventional multiclass SVM models to validate our model's performance.

ANN is designed as a three-layer network whose learning rate and momentum rate are 0.1. The hidden and output nodes use the sigmoid transfer function. We applied ANN models by varying the number of their hidden nodes from 7 to 28. Among the results, we chose the model whose number of hidden nodes is 21 since its performance was the best. This study allows 150 learning epochs for ANN. For ANN models, we applied Ward System Group's Neuroshell 4.0.

For the case of SVM-based models, the linear function, the polynomial function, and the Gaussian radial basis function were used as the kernel functions of SVM. Tay and Cao [2001] suggested that the upper bound C and kernel parameters play an important role in the performance of SVMs [Kim, 2003]. Improper selection of these two parameters can cause the overfitting or the underfitting problems. Since there is few general guidance to determine the parameters of SVM, this study varies the parameters to select the optimal values for the best prediction performance. For the case of radial basis function, Tay and Cao [2001] suggested that an appropriate range for σ^2, the bandwidth of the radial basis function, is between 1 and 100, and for C, is between 10 and 100 in their applications. Therefore, we had set the values with σ^2 as 1, 25, 50, 75, and 100 and with C, 10, 33, 55, 78, and 100. For the case of polynomial function and linear function, there is no general guideline. Thus, we had set the values of C as 10, 33, 55, 78, and 100, which are equal to the case of RBF. In case of d, the degree of the polynomial function, we took a range from 1～5.

This study used LIBSVM software system for binary classification and BSVM for multi-class classification [Chang and Lin, 2001; Hsu and Lin, 2006]. In addition, we developed a post-processing software that combines the results generated from multiple binary SVMs for the implementation of One-Against-All, One-Against-One, One-Against-The-Next, and One-Against-Followers.

For the case of MDA, we used SPSS 16.0. In addition, we employed our own software, written in Microsoft Visual Basic for Applications for Excel 2003, to implement CBR. The number of the nearest neighbor for CBR was fixed at one.

3.3 Experimental Results

We generated the hit ratios of each model, including the ordinal multiclass SVMs, the conventional multiclass SVMs, MDA, ANN, and CBR, to compare the performance of each algorithm. The hit-ratio means the ratio of the corrected cases over all the cases. The ratio is represented as follows:

$CR = \frac{1}{n}\sum_{k=1}^{n} CA_i$; CA_i = 1 if $PO_i = AO_i$, CA_i = 0 otherwise, where CR is the rate of classification accuracy for the current test-set, CA_i is a code for hit or non-hit of the i_{th} case of the test-set denoted by either 1 or 0 (“correct” = 1, “incorrect” = 0), PO_i is the predicted outcome for the i_{th} case, and AO_i is the actual outcome for the i_{th} case. The hit ratios of the ordinal multiclass SVMs and the comparative models are summarized in <Table 4>.

<Table 4> presents the predictive performances of the models evaluated. The ordinal

<Table 4> Classification Accuracies of the Ordinal Multiclass SVMs and Comparative Algorithms

<table>
<tr><th colspan="2">Model</th><th>Train set</th><th>Test set</th><th>Hold-out set</th><th>Condition</th></tr>
<tr><td colspan="2">MDA</td><td colspan="2">85.81%</td><td>87.38%</td><td>Stepwise selection</td></tr>
<tr><td colspan="2">ANN</td><td>98.04%</td><td>97.50%</td><td>96.75%</td><td># of hidden nodes = 21</td></tr>
<tr><td colspan="2">CBR</td><td colspan="2">N/A</td><td>91.88%</td><td>$k = 1$</td></tr>
<tr><td rowspan="3">Conventional Multiclass SVM</td><td>Weston and Watkins</td><td colspan="2">98.59%</td><td>97.00%</td><td>RBF*, $C = 100$, $\sigma^2 = 1$</td></tr>
<tr><td>One-Against-One</td><td colspan="2">98.69%</td><td>97.25%</td><td>RBF, $C = 100$, $\sigma^2 = 1$</td></tr>
<tr><td>One-Against-All</td><td colspan="2">98.47%</td><td>97.13%</td><td>RBF, $C = 100$, $\sigma^2 = 1$</td></tr>
<tr><td rowspan="4">Ordinal Multiclass SVM</td><td>One-Against-The-Next + Forward</td><td colspan="2">98.69%</td><td>97.25%</td><td>RBF, $C = 100$, $\sigma^2 = 1$</td></tr>
<tr><td>One-Against-The-Next + Backward</td><td colspan="2">98.69%</td><td>97.25%</td><td>RBF, $C = 100$, $\sigma^2 = 1$</td></tr>
<tr><td>One-Against-Followers + Forward</td><td colspan="2">98.69%</td><td>97.25%</td><td>RBF, $C = 100$, $\sigma^2 = 1$</td></tr>
<tr><td>One-Against-Followers + Backward</td><td colspan="2">98.69%</td><td>97.25%</td><td>RBF, $C = 100$, $\sigma^2 = 1$</td></tr>
</table>

Note) * Radial Basis Function.

multiclass SVMs were the best among the comparative models. In addition, we found that they outperformed MDA, CBR, and ANN, as well as the conventional multiclass SVMs with Weston and Watkins and One-Against-ALL. The conventional multiclass SVM with One-Against-One, however, produced the same prediction performance as the ordinal multiclass SVMs. In addition, we could not find differences among several ordinal multiclass SVMs in prediction performance. Moreover, we could not find significant statistical differences among the ordinal multiclass SVM and other comparative models except ANN and CBR.

<Table 5> Classification Results for Hold-out Samples of CBR and Ordinal Multiclass SVMs

(a) CBR ($k=1$)

		Actual Levels			
		VIP	Gold	Silver	Bronze
Predicted Levels	VIP	189	9		
	Gold	11	186	9	
	Silver		5	175	15
	Bronze			16	185

(b) 4 Types of Ordinal Multiclass SVMs

		Actual Levels			
		VIP	Gold	Silver	Bronze
Predicted Levels	VIP	200	2		
	Gold		197	4	
	Silver		1	185	4
	Bronze			11	196

Other than classification accuracy, we may also use misclassification error for each class as a criterion for evaluating the effectiveness of the proposed model. Especially, it is less critical in our case to classify low level customers into the high because they would not churn in this case. However, the opposite case is serious. In this case, the company may lose their profitable customers because of handling them as unprofitable ones. <Table 5> shows the classification results of CBR and the proposed model. In <Table 5>, the frequencies of the lower triangular area (gray area) represent the important cases for the company, which classify high-level customers into the low. The total number of these critically misclassified cases for each model is 32 (4%) and 12 (1.5%), respectively. Moreover, our proposed model predicted VIP customers with 100% accuracy. Thus, it may be another evidence that supports the value of our proposed model.

Ⅳ. Conclusions

In this study, we apply a novel multiclass SVM algorithm, the ordinal multiclass SVM, to a real life multiclass customer classification case. The experimental results show that the ordinal multiclass SVM may perform better than other traditional multiclass classification algorithms, including MDA, CBR, ANN, and most other conventional (nominal) multiclass SVM algorithms, such as the method by Weston and Watkins and One-Against-ALL from the perspective of classification performance. Moreover, our study shows that the ordinal multiclass SVMs may improve prediction results with fewer classifiers. Although One-Against-One, which is one of the nominal multiclass SVMs, produced the same prediction performance as the ordinal multiclass SVMs, the efficiency of the model might be low, because One-Against-One used more classifiers than the ordinal multiclass SVMs. We also found that the hit ratios of several ordinal multiclass SVMs were the same in this case. Thus, we can conclude that the ordinal multiclass SVMs may improve classification performance with fewer binary classifiers than the traditional classification algorithms and the conventional (nominal) multiclass SVMs vis-a-vis ordinal multiclass classification problems. We believe that there may be significant difference among the hit ratios of ordinal multiclass SVMs and other conventional multiclass models if more data are provided.

Although we compared multiple classifiers for the purpose of validating our proposed model, we have not tested certain multiclass classifiers, including multinomial logistic regression and DAGSVM (the directed acyclic graph SVM), which has been one of the most popular multiclass SVMs. In addition, in order to validate and prove the usefulness of the ordinal multiclass SVMs, there may be a need to apply our proposed model to other domains. These remain the subjects to be investigated in future research.

〈References〉

[1] Ahn, H.C. and Kim, K.-J., "Multiclass SVM model with order information," *International Journal of Fuzzy Logic and Intelligent Systems*, Vol. 6, No. 4, 2006, pp. 331-334.

[2] Ahn, H.C., Kim, K.-J., and Han, I., "Hybrid genetic algorithms and case-based reasoning systems for customer classification," *Expert Systems*, Vol. 23, No. 3, 2006, pp. 127-144.

[3] Ahn, H.C., Kim, K.-J., and Han, I., "A case-based reasoning system with the two-dimensional reduction technique for customer classification," *Expert Systems with Applications*, Vol. 32, 2007, pp. 1011-1019.

[4] Bounds, D. and Ross, D., "Forecasting customer response with neural networks," *Handbook of neural computation*, 1997, pp. 1-7.

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

[6] Coenen, F., Swinnen, G., Vanhoof, K. and Wets, G., "The improvement of response modeling: combining rule-induction and case-based reasoning," *Expert Systems with Applications*, Vol. 18, 2000, pp. 307-313.

[7] Friedman, J., "Another approach to polychromous classification," *Technical Report*, Stanford Univ., 1996.

[8] 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, 2002a, pp. 415-425.

[9] Hsu, C-W. and Lin, C.-J., "A simple decomposition method for support vector machines," *Machine Learning*, Vol. 46, 2002b, pp. 291-314.

[10] Hsu, C-W. and Lin, C.-J., "BSVM : a SVM library for the solution of large classification and regression problems," Software available at http://www.csie.ntu.edu.tw/~cjlin/bsvm/, 2006.

[11] Kim, D., Lee, H. and Cho, S., "Response modeling with support vector regression," *Expert Systems with Applications*, Vol. 34,

2008, pp. 1102-1108.

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

[13] Kreβel, U., "Pairwise classification and support vector machines," In Schölkopf, B., Burges, C. and Smola, A.J.: *Advances in Kernel Methods: Support Vector Learning*, Chapter 15. MIT Press. Cambridge, MA, 1999, pp. 255-268.

[14] Kwon, Y.S., Han, I. and Lee, K.C., "Ordinal pairwise partitioning (OPP) approach to neural networks training in bond rating," *Intelligent Systems in Accounting, Finance and Management*, Vol. 6, 1997, pp. 23-40.

[15] Lee, H. and Cho, S., "Focusing on non-respondents: Response modeling with novelty detectors," *Expert Systems with Applications*, Vol. 33, 2007, pp. 522-530.

[16] Lorena, A.C. and de Carvalho, A.C.P.L.F., "Investigation of strategies for the generation of multiclass support vector machines," In Nguyen, N.T., Katarzyniak, R.: *New Challenges in Applied Intelligence Techniques*, Springer-Verlag, Berlin, Germany, 2008, pp. 319-328

[17] Moutinho, L., Curry, B., Davies, F., and Rita, P., *Neural Network in Marketing*, Routledge, New York, 1994.

[18] Platt, J., "Fast training of support vector machines using sequential minimal optimization," In Schölkopf, B., Burges, C. and Smola, A. (ed.), *Advances in Kernel Methods-Support Vector Learning*, MIT Press, Cambridge, MA, USA, 1999.

[19] Potharst, R., Kaymak, U., and Pijls W., "Neural networks for target selection in direct marketing," Erasmus University Rotterdam in its series Discussion Paper with number 77, http://ideas.repec.org/s/dgr/eureri.html, 2001.

[20] Shin, H.J. and Cho, S., "Response modeling with support vector machines," *Expert Systems with Applications*, Vol. 30, 2006, pp. 746-760.

[21] Statnikov, A., Aliferis, C.F., Tsamardinos, I., Hardin, D., and Levy, S., "A domprehensive evaluation of multicategory classification methods for Mmicroarray gene expression cancer diagnosis," *Bioinformatics*, Vol. 21, No. 5, 2005, pp. 631-543.

[22] Suh, E.H., Noh, K.C., and Suh, C.K., "Customer list segmentation using the combined response model," *Expert Systems with Applications*, Vol. 17, No. 2, 1999, pp. 89-97.

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

[24] Vapnik, V., *The Nature of Statistical Learning Theory*, New York, NY: Springer-Verlag, 1995.

[25] Viaene, S., Baesens, B., Van Gestel, T., Suykens, J.A.K., Van den Poel, D., Vathienen, J., De Moor, B., and Dedene, G., "Knowledge discovery in a direct marketing case using least squares support vector machines," *International Journal of Intelligent Systems*, Vol. 16, 2001, pp. 1023-1036.

[26] Weston, J. and Watkins, C., "Support vector machines for multiclass pattern recognition," *Proceedings of the Seventh European Symposium on Artificial Neural Networks*, 1999, pp. 219-224.

[27] Wu, Y.-C., Lee, Y.-S. and Yang, J.-C., "Robust and efficient multiclass SVM models for phrase pattern recognition," *Pattern Recognition*, Vol. 41, No. 9, 2008, pp. 2874-2889.

◆ About the Authors ◆

Kyoung-jae Kim
Dr. Kyoung-jae Kim is an associate professor of Dept. of MIS at Dongguk University. He received his Ph.D. from KAIST. He has published in Annals of Operations Research, Applied Intelligence, Applied Soft Computing, Asia Pacific Journal of Information Systems, Computers in Human Behavior, Expert Systems, Expert Systems with Applications, Intelligent Data Analysis, Intelligent Systems in Accounting, Finance and Management, Neural Computing and Applications, Neurocomputing, and other journals. His research interests include data mining, knowledge management and intelligent agents.

Hyunchul Ahn
Dr. Hyunchul Ahn is a full-time lecturer of the School of Management Information Systems at Kookmin University in South Korea. His research interest is business applications of artificial intelligence. Dr. Ahn has published more than 30 research papers in refereed journals such as Annals of OR, Applied Soft Computing, Expert Systems and Expert Systems with Applications.