Combining Pairwise SVM Classifiers for Bond Rating

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

[a] *Graduate School of Management, KAIST*
207-43 Cheongrangri-dong, Dongdaemun-gu, Seoul 130-722, Korea
Tel: +82-2-958-3685, Fax: +82-2-958-3604, E-mail:hcahn@kaist.ac.kr, ighan@kgsm.kaist.ac.kr

[b] *Department of Information Systems, Dongguk University*
3-26 Pil-dong, Chung-gu, Seoul 100-715, Korea
Tel: +82-2-2260-3324, Fax: +82-2-2260-3684, E-mail:kjkim@dongguk.edu

Abstract

Prior studies on financial applications of data mining techniques mainly focused on the applicability of artificial neural networks, but recent research tended to investigate applicability of novel data mining techniques just like support vector machines. Support vector machines (SVMs) were originally devised for binary classification. But, in reality, there exist many problems which cannot be solved by just binary classification models, those are multiclass classification problems including bond rating. Researchers have tried to extend original SVM to multiclass classification. However, their studies have only focused on classifying samples into nominal categories. This study proposes Ordinal Multiclass SVMs which apply ordinal pairwise partitioning (OPP) to conventional SVMs in order to handle ordinal multiple classes efficiently and effectively. Our suggested model may use fewer classifiers but predict more accurately because it utilizes additional hidden information, the order of the classes. To validate our model, we apply it to the real-world bond rating case. In this study, we compare the results of the model to those of other multiclass classification algorithms. The result shows that Ordinal Multiclass SVMs improve prediction performance.

Keywords:

Support Vector Machines; Ordinal Pairwise Partitioning; Multiclass Classification; Bond Rating

Introduction

Bond ratings are important determinants of risk premiums and the marketability of bonds. But, in spite of their importance, bond ratings are typically very costly to obtain, since they require professional agencies to invest large amount of time and human resources to perform deep analysis of the company's risk status based on various aspects ranging from strategic competitiveness to operational level details. So, it has been a popular research topic for researchers to predict companies' credit ratings by applying statistical and artificial intelligence techniques (Huang et al., 2004).
Recently, support vector machines (SVMs) become popular as a solution for prediction problems because of their robustness and high accuracy. But, SVMs were originally devised for binary classification, so it doesn't fit exactly to multiclass classification just like bond rating. Researchers have tried to extend original SVM to multiclass classification. Up to now, there are two types of approaches for multiclass SVM. One is by constructing and combining several binary classifiers while the other is by directly considering all data in one optimization formulation. However, all of these approaches have only focused on classifying samples into nominal categories (Hsu and Lin, 2002a; Statnikov et al., 2005).
The aim of this study is to suggest a novel multiclass SVM approach which can handle ordinal multiple classes, such as bond ratings and customers' profitability levels, efficiently and effectively. For that purpose, we apply ordinal pairwise partitioning (OPP) approach. The OPP approach partitions the data set into sub data sets with reduced classes in the ordinal and pairwise manner according to output classes (Kwon, Han and Lee, 1996). It utilizes additional hidden information, the order of the classes, for the classification. So, it is possible to get the more accurate prediction results with fewer classifiers. To validate our model, we apply it to the real-world bond rating case. In addition, we compare the results of the model to those of multiple discriminant analysis (MDA), multinomial logistic regression (MLOGIT), case-based reasoning (CBR), artificial neural networks (ANN), and another multiclass SVM algorithm.

Conventional and Multiclass Support Vector Machines

Bond Rating Using Data Mining Techniques

Substantial studies on bond rating prediction using data mining techniques can be found in data mining literatures. The studies can be categorized three stages. The early days of these studies mainly focused on applicability of statistical techniques such as multiple discriminant analysis and logistic regression analysis. The second stage of the research is application of typical artificial intelligence techniques such as artificial neural networks (especially backpropagation networks) and case-based reasoning. Table 1 shows major prior research using artificial intelligence techniques.

Table 1 – Prior studies on bond rating using AI techniques

Study	**AI techniques**	**Benchmark methods**
Kim (1993)	BP, RBS	LinR, MDA, LogR
Moody and Utans (1995)	BP	N/A
Maher and Sen (1997)	BP	LogR, MDA
Kwon et al. (1997)	BP	MDA
Chaveesuk et al. (1999)	BP, RBF, LVQ	LogR
Shin and Han (2001)	CBR	MDA
Huang et al. (2004)	SVM	LogR, BP

BP: Backpropagation neural networks, RBS: Rule-based systems, LinR: Linear Regression, LogR: Logistic regression, RBF: Radial basis function, LVQ: Learning vector quantization, MDA: Multiple discriminant analysis

As shown in Table 1, most prior research used backpropagation neural networks. However, they suffer from difficulty in selecting a large number of controlling parameters which include relevant input variables, hidden layer size, learning rate, and momentum term.

Conventional (binary) Support Vector Machines

Conventional SVM uses linear model to implement nonlinear class boundaries through some nonlinear mapping the input vectors into the high-dimensional feature space. A linear model constructed in the new space can represent a nonlinear decision boundary in the original space. In the new space, an optimal separating hyperplane is constructed (Vapnik, 1995).
Thus, SVM is known as the algorithm that finds a special kind of linear model, the *maximum margin hyperplane*. The maximum margin hyperplane gives the maximum separation between the decision classes. The training examples that are closest to the maximum margin hyperplane are called *support vectors*. All other training examples are irrelevant for defining the binary class boundaries.
SVM constructs linear model to implement nonlinear class boundaries through the transforming the inputs into the high-dimensional feature space. The *kernel function* does this work. There are some different kernels for generating the inner products to construct machines with different types of nonlinear decision surfaces in the input space. Choosing among different kernels the model that minimizes the estimate, one chooses the best model. Common examples of the kernel function are the linear, polynomial and the Gaussian radial basis function (Tay and Cao, 2001).

Multiclass Support Vector Machines

SVMs were originally designed for binary classification, which has only one classifier. So, how to effectively extend it for multiclsss classification is still ongoing research issue. Currently, there are two types of approaches for multiclass SVM. One is by constructing several binary classifiers while the other is by directly considering all data in one optimization formulation. In detail, each approach can be classified into several methods as follows:

Constructing Several Binary Classifiers: 1-Against-All
This is conceptually the simplest multiclass method. This method constructs k binary SVM classifiers for k-class classification: class 1 versus all other classes, class 2 versus all other classes, … , class k versus all other classes (Kressel, 1999).
The combined 1-Against-All decision function chooses the class of a sample that corresponds to the maximum value of k binary classification functions specified by the furthest positive hyperplane. By doing so, the decision hyperplanes calculated by k SVMs shift, which questions the optimality of the multiclass classification (Statnikov et al., 2005).

Constructing Several Binary Classifiers: 1-Against-1
In this method, the model constructs binary SVM classifiers for all pairs of classes; in total there are ${}_kC_2$ pairs. That is, for every pair of classes, a binary SVM problem is solved with the underlying optimization problem to maximize the margin between two classes. The decision function assigns an instance to a class which has the largest number of votes, so-called *Max Wins* strategy. If ties occur, a sample will be assigned a label based on the classification provided by the furthest hyperplane (Friedman, 1996).

Constructing Several Binary Classifiers: DAGSVM
The third algorithm for constructing several binary classifiers is the directed acyclic graph SVM (DAGSVM). The training phase of this algorithm is similar to the 1-Against-1 method using multiple binary classifiers; however the testing phase of DAGSVM requires construction of a rooted binary decision directed acyclic graph (DDAG) using ${}_kC_2$ classifiers. Each node of this graph is a binary SVM for a pair of classes, say (p, q). On the topologically lowest level, there are k leaves corresponding to k classification decisions. Every non-leaf node (p, q) has two edges – the left edge corresponds to decision "not p" and the right one corresponds to "not q". An advantage of using a DAG is that some analysis of generalization can be established. There are still no similar theoretical results for above two methods yet (1-Against-All and 1-Against-1). In addition, its testing time is less than the 1-Against-1 method (Platt, Cristianini and Shawe-Taylor, 2000).

Directly Considering All Data At Once: Method by Weston and Watkins
This approach may be interpreted as a natural extension of the binary SVM classification problem. Here, in the k-class case, one has to solve single quadratic optimization problem of size $(k\text{-}l)n$ which is identical to a binary SVM for the case of k=2 (Weston and Watkins, 1999). In a slightly different formulation of QP problem, a bounded formulation, decomposition technique can provide a significant speed-up in the solution of the optimization problem (Hsu and Lin, 2002a; Hsu and Lin, 2002b).

Directly Considering All Data At Once: Method by Crammer and Singer

This method is similar to the previous one, the method by Weston and Watkins. It requires solution of a single quadratic programming problem of size *(k-l)n*, however uses less slack variables in the constraints of the optimization problem (Crammer and Singer, 2000). Similar to the method by Weston and Watkins, the use of decompositions can provide a significant speed-up in the solution of the optimization problem.

Ordinal Multiclass SVMs

As indicated in previous section, there exist several approaches to extend binary SVMs to multiclass SVMs. However, all of these methods have a common shortcoming. That is all of them were designed for the multiclass classification problems whose classes are nominal. But, in real world, there are also multiclass classification problems whose classes are ordinal (i.e. classes have orders).

In this study, we suggest Ordinal Multiclass SVM model which uses ordinal pairwise partitioning (OPP) approach as a tool for upgrading conventional multiclass SVM models in order to deal with ordinal classes wisely. The OPP approach partitions the data set into subdata sets with reduced classes in the ordinal and pairwise manner according to the output classes.

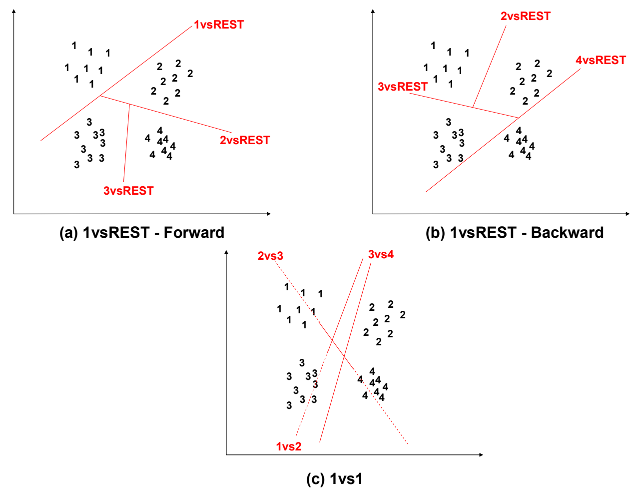


Figure 1 - Ordinal Multiclass SVMs

In general, there are two partitioning methods and two chaining methods, forward and backward. The first partitioning method is the (N-1) & N style, where N is 2, 3, 4 in the case of 4-class classification, in the forward and backward methods, called '1vs1 forward' and '1vs1 backward'. In our bond rating application, we have four bond classes. That is, we pair them to make three separate data sets, (1 & 2), (2 & 3), and (3 & 4) classes. The other partitioning method is the (N) & (remaining classes) style, which we call the '1vsREST' approach. In the 1vsREST approach, using forward and backward computation, the data set is partitioned in advance. When using the forward method, the data set is partitioned into such as a pair of (1, 2&3&4), a pair of (2, 3&4), and a pair of (3, 4). This is '1vsREST forward'. When using the backward method, the data set is partitioned to make a pair of (4, 3&2&1), a pair of (3, 2&1), and a pair of (2, 1). This is '1vsREST backward'. As we can see here, any kind of OPP approach requires just *k*-1 binary SVM models to classify data into *k* classes although other conventional requires from *k* to *k*(*k*-1)/2 models. Figure 1 shows differences between Ordinal Multiclass SVM models, 1vs1 and 1vsREST.

Experimental design and results

Research data

To validate our model, we apply it to the real world bond rating data from National Information and Credit Evaluation, Inc., a major bond rating company in Korea. We obtained the bond rating results for the year 2002 and various financial variables from 1,295 companies in manufacturing industry in Korea. In Korean bond rating market, bond ratings are divided into 5 classes, A1, A2, A3, B and C. But, we adjust our data to 4 classes by combining B and C ratings into one group because their numbers of samples were so small. And both ratings are usually treated same as just junk bonds in the market.

Table 2 - Definition of selected variables

Variables	Definitions
SHEQ	Shareholder's equity
SALE	Sales
DEBT	Total debt
SAPE	Sales per employee
NIPS	Net income per share
YEAR	Years after founded
AETA	Accumulated earning to total asset
BDRA	Borrowings-dependency ratio
FCTC	Financing cost to total cost
FIRA	Fixed ratio
IACA	Inventory assets to current assets
SBTB	Short-term borrowings to total borrowings
CFTA	Cash flow to total assets
OACF	Cash flows from operating activity

Original data consists of 39 financial ratio variables that are known to the features affecting bond rating in previous literatures. Among them, we select 36 variables by applying two-samples T-tests and, finally, select 14 variables which are proved to be the most influential in bond rating by applying stepwise statistical method. The selected variables are presented in Table 2. In this study, 20% of the data for each class are used for validation and the remaining 80% of data were used for training. And, to overcome the scarcity of samples, we adopt 5-fold cross-validation.

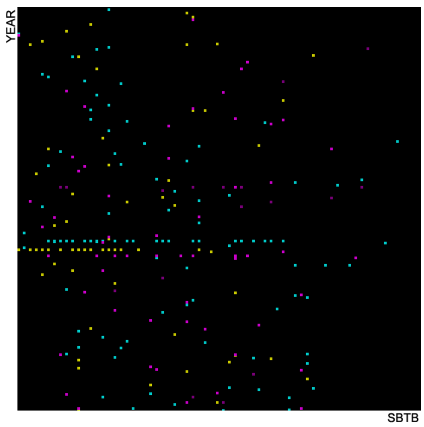
Experimental design

To validate the superiority of our models' performances with sophistication, we apply our Ordinal Multiclass SVM

models as well as MDA, multinomial logistic regression (MLOGIT), CBR, ANN, and multiclass SVM models by the method of Crammer and Singer (C&S).
In the case of ANNs, we adopt standard three-layer back propagation networks and set the number of nodes in the hidden layer as 7, 14, 21 and 28. For the stopping criteria of ANNs, this study allows 50 learning epochs and set the learning rate to 0.1 and the momentum term to 0.1. The hidden and output nodes use the sigmoid transfer function.
In the case of SVM-based models, the linear function, the polynomial function and the Gaussian radial basis function are used as the kernel function of SVM. Tay and Cao (2001) showed that the upper bound *C* and the kernel parameters play an important role in the performance of SVMs. 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 optimal values for the best prediction performance. This study uses LIBSVM software system for binary classification and BSVM for multiclass classification by the method of C&S[1].

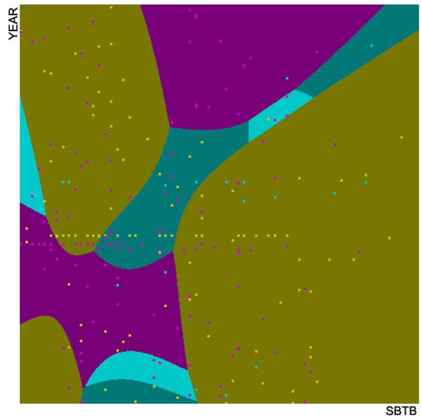
Experimental results

To compare the performance of each algorithm, we calculate the hit ratios of them. The hit ratios of the Ordinal Multiclass SVMs and the comparative algorithms are summarized in Table 3. Table 3 shows that the performance of 1vsREST-Forward algorithm is the best among all the types of Ordinal Multiclass SVMs. In addition, we can find that it outperforms MDA, MLOGIT, CBR, ANN as well as the multiclass SVM method by C&S.
Figure 2 presents the results of two multiclass SVM models for the validation data set of the first cross validation set. As we can see from Table 3, 1vsREST-Forward method for Ordinal Multiclass SVM shows the best performance. Figure 2(a) represents data patterns before SVM is employed and 2(b) represents the best results of the method by C&S. And, Figure 2(c) shows the best result of the Ordinal Multiclass SVM model. All of these are created by modified version of 'svmtoy' and 'bsvmtoy', included tools in LIBSVM and BSVM.

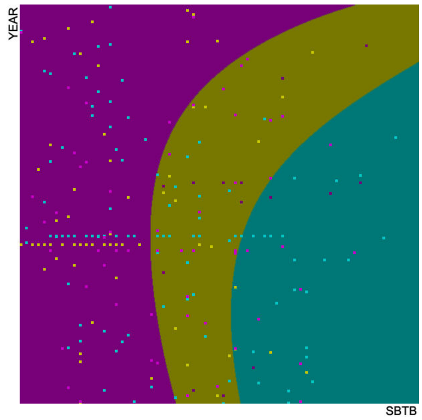


(a) The data patterns before SVM is employed

[1] BSVM is available at http://www.csie.ntu.edu.tw/~cjlin/bsvm/. LIBSVM is available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/.



(b) The best result of the Crammer & Singer's method



(c) The best result of the Ordinal Multiclass SVM

Figure 2 – Graphical results of multiclass SVM models

The McNemar tests are used to examine whether the predictive performance of the Ordinal Multiclass SVM is significantly higher than that of other algorithms. This test is used with nominal data and is particularly useful with before-after measurement of the same subjects (Kim, 2003). Table 4 shows the results of the McNemar test to compare the performances of five algorithms for the test data.
As shown in Table 4, Ordinal Multiclass SVM (1vsREST – Forward) is better than MDA and CBR at the 1%, and better than MLOGIT, ANN and the method by C&S at the 5% statistical significance level.

Conclusions

In this study, we propose a new multiclass SVM algorithm, the Ordinal Multiclass SVM, and apply it to the bond rating case. The experimental results show that the Ordinal Multiclass SVM may result in better performance than other traditional multiclass classification algorithms including MDA, MLOGIT, CBR, ANN as well as another multiclass SVM algorithm, the method by Crammer & Singer, from the perspective of classification performance. Moreover, our study shows the Ordinal Multiclass SVMs may improve the prediction results with fewer classifiers.
However, in order to validate and prove its usefulness, it is required to apply the proposed model to other domains.

Table 3 - The classification accuracies of Ordinal Multiclass SVMs and all comparative algorithms

Data Sets	MDA	MLOGIT	CBR	ANN	C&S	Ordinal Multiclass SVM			
						1vs1		1vsREST	
						Forward	Backward	Forward	Backward
1	59.30%	63.95%	48.45%	64.73%	65.12%	65.50%	65.50%	66.28%	65.89%
2	62.02%	63.57%	47.67%	65.12%	62.79%	64.34%	65.12%	65.12%	64.73%
3	69.38%	68.99%	54.26%	66.67%	64.73%	69.77%	69.38%	71.32%	70.93%
4	65.50%	67.44%	56.98%	67.44%	66.67%	68.22%	68.22%	68.60%	68.60%
5	59.30%	63.18%	49.61%	64.34%	65.12%	68.99%	67.44%	68.60%	67.83%
Avg.	63.10%	65.43%	51.40%	65.66%	64.89%	67.36%	67.13%	67.98%	67.60%

Table 4 - McNemar values for the hold-out data

	MLOGIT	CBR	ANN	C&S	Ordinal MultiSVM
MDA	4.163*	49.020**	5.044*	1.238	15.315**
MLOGIT		73.469**	0.028	0.131	5.198*
CBR			76.111**	68.331**	109.587**
ANN				0.285	3.823*
C&S					5.394*

* significant at the 5% level, ** significant at the 1% level

References

[1] Chaveesuk, R., Srivaree-Ratana, C. and Smith, A.E. (1999). "Alternative Neural Network Approaches to Corporate Bond Rating," *Journal of Engineering Valuation and Cost Analysis*, Vol. 2, pp. 117-131.

[2] Crammer, K. and Singer, Y. (2000). "On the Learnability and Design of Output Codes for Multiclass Problems," *Proceedings of the 13th Annual Conference on Computational Learning Theory*, pp. 35-46.

[3] Friedman, J. (1996). "Another Approach to Polychtomous Classfication," Technical Report, Stanford Univ.

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

[5] Hsu, C-W. and Lin, C.-J. (2002b). "A Simple Decomposition Method for Support Vector Machines," *Machine Learning*, Vol. 46, pp. 291-314.

[6] Huang, Z., Chen, H., Hsu, C.-J., Chen, W-H. and Wu, S. (2004). "Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study," *Decision Support Systems*, Vol. 27, pp. 543-558.

[7] Kim, J.W. (1993). "Expert Systems for Bond Rating: A Comparative Analysis of Statistical, Rule-based and Neural Networks Systems," *Expert Systems*, Vol. 10, pp. 167-171.

[8] Kim, K. (2003). "Financial Time Series Forecasting using Support Vector Machines," *Neurocomputing*, Vol. 55 (1-2), pp. 307-319.

[9] Kressel, U. (1999). "Pairwise Classification and Support Vector Machines," In Schölkopf, B., Burges, C. and Smola, A.J.: *Advances in Kernal Methods: Support Vector Learning*, Chapter 15. MIT Press. Cambridge, MA, pp. 255-268.

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

[11] 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*, Vol. 6, pp. 59-72.

[12] Moody, J. and Utans, J. (1995). "Architecture Selection Strategies for Neural Network Application to Corporate Bond Rating," In Refenes, A.: *Neural Networks in the Capital Markets*. Wiley. Chichester, pp. 277-300.

[13] Platt, J.C., Cristianini, N. and Shawe-Taylor, J. (2000). "Large Margin DAG's for multiclass classification," In Solla, S.A., Leen, T.K. and Muller, K.-R.: *Advances in Neural Information Processing Systems*, Vol. 12. MIT Press. Cambridge, MA, pp. 547-553.

[14] Shin, K.S. and Han, I. (2001). "A Case-based Approach using Inductive Indexing for Corporate Bond Rating," *Decision Support Systems*, Vol. 32, pp. 41-52.

[15] Statnikov, A., Aliferis, C.F., Tsamardinos, I., Hardin, D. and Levy, S. (2005). "A Comprehensive Evaluation of Multicategory Classification Methods for Microarray Gene Expression Cancer Diagnosis," *Bioinformatics*, Vol. 21(5), pp. 631-543.

[16] Tay, F.E.H. and Cao, L.J. (2001). "Application of Support Vector Machines in Financial Time Series Forecasting," *Omega*, Vol. 29, pp. 309-317.

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

[18] Weston, J. and Watkins, C. (1999). "Support Vector Machines for Multiclass Pattern Recognition," *Proceedings of the Seventh European Symposium on Artificial Neural Networks*, pp. 219-224.