International Journal of Fuzzy Logic and Intelligent Systems, vol. 6, no. 4, December 2006 pp. 331-334

Multiclass SVM Model with Order Information

Hyunchul Ahn*, Kyoung-jae Kim**

*** Graduate School of Management, Korea Advanced Institute of Science & Technology**
**** Department of MIS, Dongguk University, Corresponding Author**

ABSTRACT

Original Support Vsctor Machines (SVMs) by Vapnik were used for binary classification problems. Some 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 a novel multiclass SVM model in order to handle ordinal multiple classes. Our suggested model may use less 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 statistical and typical machine learning techniques, and another multiclass SVM algorithm. The result shows that proposed model may improve classification performance in comparison to other typical multiclass classification algorithms.

Key Words : Multiclass SVM, Order information, Bond rating

1. Introduction

Multiclass support vector machines (Multiclass SVMs) were devised for multiclass classification problems because of the fact that there exist many problems which cannot be solved by just binary classification models. 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 [1,3,4,5,7,10,12,16].

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 to multiclass SVM. The OPP approach partitions the data set into sub data sets with reduced classes in the ordinal and pairwise manner according to output classes [8]. These approaches utilize additional hidden information, the order of the classes, for the classification. So, it is possible to get the more accurate prediction results by using less 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 multinomial logistic regression, artificial neural networks, and other multiclass SVM algorithms.

The rest of this paper is organized as follows: The next section reviews conventional multiclass support vector machines and proposes a novel multiclass SVM algorithm. Section 3 describes the data and the experiments. In that section, the empirical results are also summarized and discussed. In the final section, conclusions and the limitations of this study are presented.

Manuscript received Sep. 13, 2006; revised Dec. 18, 2006.

II. OPP Approach to Multiclass SVMs

SVMs have become an increasingly popular technique for construct classificarion models. SVM uses a linear model to implement nonlinear class boundaries by nonlinear mapping of the input vectors x into the high-dimensional feature space. A linear model constructed in the new space can represent a nonlinear boundary in the original space. In the new space, an optimal separating hyperplane is constructed [14,15].

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 a linear model to implement nonlinear class boundaries through the transformation of the inputs into the high-dimensional feature space. The function, K(xi,xj), which is called '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 polynomial kernel $K(xi,xj)= (1+ xi^{T}xj)^{d}$ and the Gaussian radial basis function (RBF) $K(xi,xj)= \exp(-1/\delta^{2}(xi - xj)^{2})$ where d is the degree of the polynomial kernel and δ^{2} is the bandwidth of the Gaussian RBF kernel.

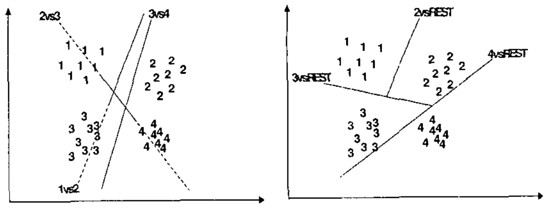
In general, there are no parameters to tune except the upper bound C for the non-separable cases in linear SVM [2]. Thus, overfitting is unlikely to occur with SVM. Overfitting may be caused by too much flexibility in the decision boundary, but the maximum hyperplane is relatively stable and gives little flexibility [17].

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. The former consists of 1-Against-All (1-A-All in this study), 1-Against-1 (1-A-1 in this study), and DAGSVM (directed acyclic graph SVM), the latter consists of the methods of Weston & Watkins and Crammer & Singer.

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 many multiclass classification problems whose classes are ordinal (i.e. classes have orders). For example, bond ratings are usually classified into five classes, AAA, AA, A, B and C. In this case, AAA is better than AA and AA is better than A, and so on. The classes of bond ratings are ordinal. So, it may be important research issue to develop a methodology for multiclass classification which can deal with ordinal classes efficiently and effectively.

In this study, we suggest 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 [8]. In general, there are two partitioning methods. One partitioning method is the (N-1) & N style, which we call the '1vs1' approach. 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. 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 $\frac{k(k-1)}{2}$ models. Fig. 1. shows differences between OPP-based SVMs, 1vs1 and 1vsREST.

After dividing data into $k-1$ groups for OPP processing, we train and evaluate $k-1$ binary SVM models with each of the three data sets above. That is, in the 1vs1 approach, we obtain SVM Model 1 for (1 & 2) classes, Model 2 for (2 & 3) classes and Model 3 for (3 & 4) classes. In the same manner, four SVM models are obtained in the 1vsREST approach. Each SVM model trained by the ordinal pairing with reduced classes of two is trained well enough to differentiate between two classes, eventually mitigating the difficulty of doing the nonlinear functional mapping [8].



(a) 1vs1 (b) 1vsREST

Fig. 1. OPP-based multiclass support vector machines applied to a four-class classification problem. (a) represents the method by 1vs1, while (b) represents the one by 1vsREST.

III. Experimental design and results

3.1. Research data and experimental design

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, AAA, AA, A, 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.

Original data consists of 39 financial ratio variables that are known to the features affecting bond rating in previous literatures [6,11]. Among them, we select 36 variables by applying two-samples T-test and, finally, select 14 variables which are proved to be the most influential in bond rating by applying stepwise statistical method. 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.

To validate the superiority of our models' performances with sophistication, we apply our OPP-based multiclass SVM models as well as multinomial logistic regression (MLOGIT), artificial neural networks (ANNs), and other multiclass SVM algorithms.

For MLOGIT, we use multinomial logistic regression module of SPSS for Windows 13.0. 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 nodes use the sigmoid transfer function and the output node uses the linear transfer function. Experiment for the ANN is conducted by Neuroshell2 4.0.

In the case of SVM-based models, the linear kernel, the polynomial kernel and the Gaussian radial basis function are used as the kernel function of SVM. Tay and Cao [13] showed that the upper bound C and the kernel parameter play an important role in the performance of SVMs. Improper selection of these two parameters can cause the overfitting or the underfitting problems. So, this study varies the parameters to select optimal values for the best prediction performance.

3.2. Experimental results

To compare the performance of each algorithm, we calculate the hit ratio of each of them. The hit ratios of the OPP-based multiclass SVMs are summarized in Table 1 and ones of the comparative algorithms are summarized in Table 2 and 3. These tables show that the performance of 1vsREST algorithm of OPP-based SVMs (OPP2 in Table 1) is the best among all comparative methods for the validation data set.

Table 1. The result of OPP-based multiclass SVMs (Hit Rate, %)

Data sets	OPP1 : 1vs1		OPP2 : 1vsREST	
	Train	Valid	Train	Valid
1	65.57	65.50	73.19	65.89
2	80.14	65.12	80.62	64.73
3	66.06	69.38	66.06	70.93
4	74.25	68.22	80.14	68.60
5	73.87	67.44	67.89	67.83
Avg	71.98	67.13	73.58	67.60

Table 2. The result of other multiclass SVMs (Hit Rate, %)

Data sets	1-Against-1 (1-A-1)		1-Against-All (1-A-All)	
	Train	Valid	Train	Valid
1	75.80	65.12	73.10	65.50
2	84.09	65.12	76.37	63.18
3	67.21	69.38	63.36	64.34
4	76.47	68.22	80.04	65.12
5	75.31	68.22	64.80	63.57
Avg	75.78	67.21	71.53	64.34

Table 3. The result of comparative algorithms (Hit Rate, %)

Data sets	MLOGIT*		ANN**		
	Train	Valid	Train	Test	Valid
1	67.70	63.95	70.22	64.73	64.73
2	67.98	63.57	71.76	65.50	65.12
3	65.96	68.99	69.45	63.18	66.67
4	67.79	67.44	69.19	64.73	67.44
5	66.15	63.18	67.14	68.60	64.34
Avg	67.12	65.43	69.55	65.35	65.66

* multinomial logistic regression,
** artificial neural networks

The McNemar tests are used to examine whether the predictive performance of the proposed model 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. Table 4 shows the results of the McNemar test to compare the performances of six algorithms for the validation (holdout) data.

As shown in Table 4, 1vsREST algorithm of the OPP-based multiclass SVM (OPP2) is better than 1-Against-All at the 1%, and MLOGIT at the 5%, and ANN at the 10% statistical significance level.

Table 4. McNemar values for the holdout data

	ANN	1-A-1	1-A-All	OPP1	OPP2
MLOGIT	0.028	2.766*	0.693	2.657	4.005**
ANN		1.861	1.020	1.662	2.756*
1-A-1			5.918**	0.000	0.147
1-A-All				5.518**	6.946***
OPP1					0.187

*** significant at the 1% level, ** significant at the 5% level,
* significant at the 10% level

IV. Concluding remarks

In this study, we propose new multiclass SVM algorithms for ordinal multiclass classification, OPP-based SVMs, and apply them to the bond rating case. The experimental results show that OPP-based SVMs may result in better performance than other traditional multiclass classification algorithms including MLOGIT, ANN as well as another multiclass SVM algorithms, the method of constructing several binary classifiers, from the perspective of hit ratio. Our study shows the OPP-based SVMs may improve the prediction results with less classifiers.

However, in order to validate and prove its usefulness, it is needed to test the performances of other multiclass SVM models.

References

1. Crammer, K., Singer, Y.: On the Learnability and Design of Output Codes for Multiclass Problems. *Proc. of the 13th Annual Conference on Computational Learning Theory* (2000) 35-46
2. Drucker, H., Wu, D., Vapnik, V.N.: Support vector machines for spam categorization. *IEEE Transactions on Neural Networks*, 10(5) (1999) 1048-1054
3. Friedman, J.: *Another Approach to Polychtomous Classfication.* Technical Report, Stanford University (1996)
4. Hsu, C-W, Lin, C.-J.: A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Transactions on Neural Networks* 13(2) (2002) 415-425
5. Hsu, C-W., Lin, C.-J.: A Simple Decomposition Method for Support Vector Machines. *Machine Learning* 46

(2002) 291-314.
6. Huang, Z., Chen, H., Hsu, C.-J., Chen, W-H., Wu, S.: Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study. *Decision Support Systems* 27 (2004) 543-558
7. Kressel, U.: Pairwise Classification and Support Vector Machines. In Schölkopf, B., Burges, C., Smola, A.J.: *Advances in Kernal Methods: Support Vector Learning* Chapter 15. MIT Press. Cambridge, MA (1999) 255-268
8. Kwon, Y.S., Han, I., Lee, K.C.: Ordinal Pairwise Partitioning (OPP) Approach to Neural Networks Training in Bond Rating. *Intelligent Systems in Accounting, Finance and Management* 6 (1997) 23-40
9. Mukherjee, S., Osuna, E., Girosi, F.: Nonlinear prediction of chaotic time series using support vector machines. *Proc. of the IEEE Workshop on Neural Networks for Signal Processing* (1997) 511-520
10. Platt, J.C., Cristianini, N., Shawe-Taylor, J.: Large Margin DAG's for multiclass classification. In Solla, S.A., Leen, T.K., Muller, K.-R.: *Advances in Neural Information Processing Systems* 12. MIT Press. Cambridge, MA (2000) 547-553
11. Shin, K.S., Han, I.: A Case-based Approach using Inductive Indexing for Corporate Bond Rating. *Decision Support Systems* 32 (2001) 41-52
12. Statnikov, A., Aliferis, C.F., Tsamardinos, I., Hardin, D., Levy, S.: A Comprehensive Evaluation of Multicategory Classification Methods for Microarray Gene Expression Cancer Diagnosis. *Bioinformatics* 21(5) (2005) 631-543
13. Tay, F.E.H, Cao, L.J.: Application of Support Vector Machines in Financial Time Series Forecasting. *Omega* 29 (2001) 309-317
14. Vapnik, V.: *The Nature of Statistical Learning Theory.* Springer-Verlag. New York (1995)
15. Vapnik, V.N.: *Statistical Learning Theory.* Wiley. New York (1998)
16. Weston, J., Watkins, C.: Support Vector Machines for Multiclass Pattern Recognition. *Proc. of the Seventh European Symposium on Artificial Neural Networks* (1999) 219-224
17. Witten, I.H., Frank, E.: *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations.* Morgan Kaufmann Publishers. San Francisco, CA (2000)

Hyunchul Ahn
Received the B.E., M.E., and Ph.D. degrees from Korea Advanced Institute of Science and Technology (KAIST). He works as a postdoctoral researcher at KAIST Business School. His research interests are in the areas of data mining in marketing and finance and artificial intelligence techniques such as case-based reasoning, genetic algorithms and support vector machines.

Phone : +82-2-958-3685
E-mail : hcahn@kaist.ac.kr

Kyoung-jae Kim
Received his B.B.A. degree from Chung-Ang University and M.E. and Ph.D. degrees in Management Information Systems from KAIST. He is currently an assistant professor in the Department of Management Information Systems, Dongguk University. He published his papers in *Applied Intelligence, Expert Systems, Expert Systems with Applications, Intelligent Data Analysis, Intelligent Systems in Accounting Finance & Management, Neural Computing & Applications, Neurocomputing*, and other journals. His research interests include data mining, knowledge management, and intelligent agents.

Phone : +82-2-2260-3324
E-mail : kjkim@dongguk.edu