Multiclass Support Vector Machines with Simultaneous Multi-Factors Optimization for Corporate Credit Ratings

Hyunchul Ahn, William X. S. Wong

***Abstract*—**Corporate credit rating prediction is one of the most important topics, which has been studied by researchers in the last decade. Over the last decade, researchers are pushing the limit to enhance the exactness of the corporate credit rating prediction model by applying several data-driven tools including statistical and artificial intelligence methods. Among them, multiclass support vector machine (MSVM) has been widely applied due to its good predictability. However, heuristics, for example, parameters of a kernel function, appropriate feature and instance subset, has become the main reason for the critics on MSVM, as they have dictate the MSVM architectural variables. This study presents a hybrid MSVM model that is intended to optimize all the parameter such as feature selection, instance selection, and kernel parameter. Our model adopts genetic algorithm (GA) to simultaneously optimize multiple heterogeneous design factors of MSVM.

***Keywords*—**Corporate credit rating prediction, feature selection, genetic algorithms, instance selection, multiclass support vector machines.

I. Introduction

CORPORATE credit rating assessment consists of complicated processes in which various factors describing a company are taken into consideration. However, the requirement for domain experts to access the rating for corporate credit rating assessment is known to be pricy. As such, researcher and practitioners have now considering the use of data-driven corporate credit rating prediction using statistical and artificial intelligence (AI) techniques to perform the said assessment. In particular, statistical methods such as multiple discriminant analysis (MDA) and multinomial logistic regression analysis (MLOGIT), and AI methods including case-based reasoning (CBR), artificial neural network (ANN), and multiclass support vector machine (MSVM) have been applied to corporate credit rating [1]-[5].

Despite of being popular for its robustness and high prediction, the application of MSVM is not simple since it requires the proper setting of some design factor such as selecting a suitable kernel function and its parameters (e.g., C, d, σ^2). The prediction performance of MSVM may be influenced by the selection of the appropriate feature subset [6]–[8]. Improvement can be done on the classification accuracy of MSVM by selecting the proper instance selection (that is, prototype selection) that will eliminate irrelevant and distorting training samples [9]. Nonetheless, these design factors were set by heuristics in most prior studies on MSVM.

Under this background, this study proposes a novel hybrid MSVM model with the simultaneous optimization of the design factors for MSVM to predict corporate credit rating prediction better. Our model, named GOMSVM, is designed to simultaneously optimize the kernel parameters, feature subset selection, and instance selection. Simultaneous optimization of multiple design factors may lead to improving the accuracy of prediction with the synergetic effect [10]–[12]. In order to optimize the multiple heterogeneous factors all at once, we present GA, which has been applied by numerous prior studies as an optimization tool.

Hyunchul Ahn is with the Graduate School of Business IT, Kookmin University, Seoul 02707 South Korea (phone: +82-2-910-4577; fax: +82-2-910-4017; e-mail: hcahn@kookmin.ac.kr).

William X. S. Wong is with the Graduate School of Business IT, Kookmin University, Seoul 02707 South Korea (e-mail: williamwong @kookmin.ac.kr).

II. Theoretical Background

A. Corporate Credit Rating Prediction

Substantial researches has been done by academics in assessing corporate credit rating since it is necessary for risk management in companies and financial institutions. In order to predict corporate credit rating, most prior studies use the signals of financial data or ratio and adopted various data-driven techniques to complete their studies. The published research can be conceptualized as evolving in three phases [3]. In the first phase, applicability of statistical techniques was the main focus in the early investigation of techniques for credit rating. For example, [13] and [14] investigated MDA, and [15] used ordered probit regression (OPR), and logistic regression analysis (LogR). Meanwhile in the second phase, application of typical techniques of AI, such as ANN and CBR was featured [16]–[21]. In particular, backpropagation neural network, (BPN), a kind of ANN, was most frequently applied [16]–[20]. However, BPN suffers from issues relating to the selection of a large number of control parameters that pertain to the relevant input variables, hidden layer size, learning rate, and momentum term. Moreover, large amount of data for the training model is required by BPN due to the constraint on degrees of freedom. To overcome these limitations, recent studies have sought to apply MSVM for corporate credit rating.

Several techniques of MSVM such as One-Against-One and method of Crammer & Singer was adopted by [5] in building the prediction models of credit rating. Different parameter was investigated in searching for the optimal MSVM model. Finally, they opted for the method of Crammer & Singer, using an RBF kernel function with σ^2=10 and C=1000. They found that this MSVM model outperformed not only BPN, but also LogR, in the prediction of bond rating for Taiwan and the U.S.

One-vs-All, One-vs-One, and DAGSVM was employed by [22] to predict the S&P's bond rating. For the kennel function, Gaussian RBF was applied and the optimal parameters of of σ^2 and C were derived from a grid-search strategy. Among the three methods, DAGSVM give the best performance. Furthermore, multiclass classification techniques, including LogR, OPR, and BPN were outdone by all types of MSVM approaches.

An automatic-classification model for credit ratings using the One-vs-One approach was built by [23] for Taiwan issuers. Similar to [22], they also adopted a Gaussian RBF kernel function and a grid-search strategy for determining optimal parameters. They found that the MSVM model was statistically superior to BPN and LogR models.

Similar to [23], [24] uses the same method to build a corporate credit-rating prediction model for Korean companies. Gaussian RBF kernel function and grid-search strategy was also applied to determine the optimal parameters. In his study, [24] found that the MSVM model had significantly outperformed BPN, MDA, and CBR.

B. Multiclass Support Vector Machine

Initially, SVMs were designed for binary classification, which has only one classifier [25] and the research is still continuing for the extension of this original SM to multiclass SVM [3]. In general, 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: One-Vs-All.** This is conceptually the simplest multiclass method. This method constructs k binary SVM classifiers for k-class classification: class 1 (positive) versus all other classes (negative), class 2 versus all other classes, …, class k versus all other classes [26].
  The combined One-vs-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. In the process, the decision hyperplanes calculated by k SVMs shift, which questions the optimality of the multiclass classification [3], [27].
- **Constructing Several Binary Classifiers: One-vs-One**. 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 [3] [26].
- **Constructing Several Binary Classifiers: DAGSVM**. The third algorithm for constructing several binary classifiers is the directed acyclic graph SVM (DAGSVM) [28]. The training phase of this algorithm is similar to the One-Against-One 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". The choice of the class order in the DDAG list can be arbitrary as shown empirically in [28].
- **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 $(k - l)n$, which is identical to binary SVMs for the case of k=2 [29]. 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 [30].
- **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 QP problem of size $(k - l)\ n$, however uses less slack variables in the constraints of the optimization problem [31]. 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 [30].

C. Genetic Algorithm

In the attempt to simulate biological evolution phenomenon, GA is well known as an efficient and effective search methods. The search result was gradually improved by applying the genetic operations. In particular, GA was avoided from falling into local optima by mutation mechanism and the search time was reduced by crossover mechanism. The general evolution process of GA proceeds as follows:

First, set of solutions known as the population was randomly created. And, each solution in the population is called a chromosome. A chromosome should be designed to represent a solution, and it is designed in most cases as the form of a binary string.

Next, popularly used genetic operators such as selection, crossover, and mutation were applied to the initial population generated. The fitness value of each chromosome, calculated from a user-defined function, allow the selection operator to evaluate and select the fittest chromosomes. The user-defined function is known as *fitness function*. Fitness function uses accuracy to classify problems. By exchanging the genes of two parent chromosomes, the crossover operator is able to obtain a new offspring to get a better solution. In the mutation operator, the bits that are arbitrarily selected with very low probability are inverted. Based on the philosophy of survival of the fittest, a new population that consists of the fittest chromosomes and the offspring of these chromosomes can be formed using the mentioned genetics operators.

Unless the stopping conditions are fulfilled, the said evolution procedure will continue to create new population using the mentioned genetic operators [12].

D. Optimization of MSVM

Despite of the various ways proposed by many researchers to optimize the design factors of SVM, it is quite rare to find any prior studies that attempt to optimize the design factor of MSVM [4]. Lorena and de Carvalho [6] are one of the pioneers that tried to combine MSVM and they also proposed a new MSVM model. This new model of MSVM uses GA to optimize the kernel parameters. They found that, when applying the proposed model to iris recognition, a better prediction accuracy was obtained.

For the sake of better classification of cell phones types, [7] suggested optimal feature subset selection approach for MSVM. RFE (recursive feature elimination) was proposed instead of GA as a tool for optimal feature selection. However, they [7] did not pay any attention on the optimization of other design factors, which include the kernel parameter, thus making their study having some limitation.

Hong and Park [31] also proposed the optimization of feature selection for MSVM, but they applied it to the corporate credit ratings prediction of the S&P 500 firms. As an approach for feature selection, they proposed to use the impurity measure of decision trees instead of RFE or GA. Similar to [7], their study ignored the optimization of the other design factors for MSVM.

Recently, a proposal regarding to simultaneous optimization of kernel parameters and feature subset selection was suggested by [8] and [4]. Since the classification performance of MSVM will be affected by both kernel parameters and feature selection, [8] and [4] think that it is more reasonable to optimize the factors all at once. These studies reported that the simultaneous optimization model leads to better prediction accuracy in rock-type classification [8] and corporate credit rating classification [4].

From a wide variety of empirical tests, a newly proposed instance selection method for MSVM by [9] demonstrated fewer instances selection and maintaining high classification accuracy on most dataset. However, their method only included the design factor optimization for instance selection and excluded the other design factors.

III. Proposed Model

In this study, we present a MSVM model, which simultaneously optimizes feature selection, instance selection, and kernel parameters by using GA. As described in Section II *D*, there have been a few prior studies that tried to optimize the design factors of MSVM. However, based on our survey, no attempts have been done to simultaneously optimize all of the design factors in MSVM using GA, in order to construct a better corporate credit rating prediction model. Therefore, a global optimization model that simultaneously optimizes the design factors of MSVM using GA for corporate credit rating prediction is proposed.

Our proposed model is named as GOMSVM, which indicates *Globally Optimized MSVM*. Fig. 1 illustrates the process that how GOMSVM works.

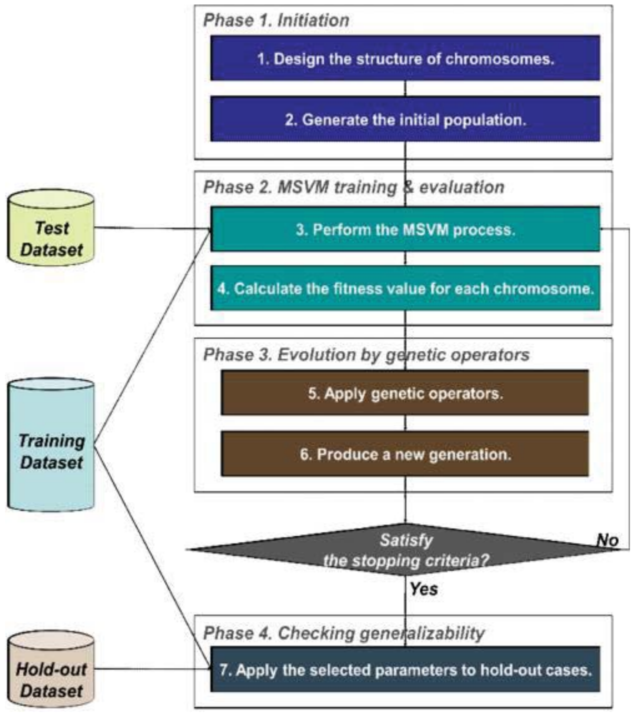

Fig. 1 Process of GOMSVM

The detailed explanation on each phase of GOMSVM is as follows:

A. Phase 1: Initiation

In order to generate the initial population of GA search needed in the first phase, the structure of a chromosome will be designed in advance. The values to be encoded into a chromosome should be transformed into binary string before the genetic operators are applied. All information regarding to feature selection, instance selection and kernel parameter settings should be included in each chromosome in GOMSVM. The selections are generally easy to be encoded as binary strings since the values of the codes for feature selection and instance selection are set to '0' or '1'. Here, '0' denotes the corresponding feature or instance is not selected and '1' denotes it is selected. In the case of kernel parameters, their values should be converted to binary numbers. Since the precision level will be manipulated by the bits assigned to the kernel parameters, GOMSVM will be using 14 bits per parameter and Gaussian RBF will be applied as the kernel function because Gaussian RBF has been proven to be the best performing kernel function in some prior studies [12]. The chromosomes of GOSVM was designed to optimize two kernel parameters (C, σ^2) of Gaussian RBF kernel function, which according to [32], these upper bound C and kernel parameter σ^2 are critical to the performance of SVM when using Gaussian RBF. Finally, the length of each chromosome becomes

m+n+28 bits, where m is the number of features, and n is the number of instances.

Initial population is generated once the chromosome structure design is done. At this time, the values of the chromosomes in the population are set to random values before the search process.

B. Phase 2: MSVM Training and Evaluation

Phase 2 involve the procedure in which GOMSVM will perform a typical MSVM process repeatedly according to the assigned value of the factors in the chromosomes. Among the several option mentioned in section II.*B*, One-vs-One approach was chosen to implement MSVM, since it is known as the most accurate approach among the others [2], [4]. GOMSVM will then evaluate the fitness value of each chromosome. Finding the optimal or near optimal design factors that lead to the most accurate prediction was the main objective of GA search in the mentioned GOMSVM. From this perspective, we use the prediction accuracy of the test data set as the fitness function of GOMSVM [4], [10]–[12].

C. Phase 3: Evolution by Genetic Operators

A new generation of population will be produced when GOMSVM applies the genetic operators, such as selection, crossover, and mutation, to the current population based on the evaluation results obtained in Phase 2.

Phase 2, which is the MSVM training process with the evaluation of fitness value, will be carrying out again after the production of a new generation in order to evaluate the newly generated population. The next generation is then created in Phase 3. That is, from this point, iteration of Phase 2 and Phase 3 will be continue until the stopping condition are fulfilled. When the stopping conditions are satisfied, the chromosome that shows the best fitness value in the last population is finally selected, and the optimal values of GOMSVM's design factors (i.e., feature selection, instance selection, and kernel parameters) are determined according to the values encoded on the chromosome.

D. Phase 4: Checking Generalizability

Guided by the prediction accuracy of the test data set, GA search of GOMSVM are able to fit the optimized design factors determined by GA into the test data. However, when applying the optimized design factor to an unknown data set, poor prediction performance was observed occasionally. Overfitting is the reason why this phenomenon happens. The design factors of GOMSVM may lose general applicability when it fit too well with the given test data set. GOMSVM will apply the final selected design factors, which is the optimal selection features and instances, and the optimal kernel parameters, to the hold-out (unknown) data set in the last phase in order to prevent overfitting and also to check for generalizability of the determined factors.

IV. Conclusion

In order to improve the performance of the typical MSVM algorithm for corporate credit rating prediction, a new kind of hybrid MSVM and GA model named GOMSVM is proposed in this study. Further, GA is proposed in this study as a tool for simultaneously optimizing multiple design factors such as feature selection, instance selection, and kernel parameters.

Validation of the usefulness and applicability of the proposed model require empirical validation using real-world data set. Also, the development of the software for the experiments is also needed. Moreover, it is necessary to compare the prediction performances of GOMSVM and other comparative models, in order to examine the superiority of GOMSVM.

References

[1] H. Ahn, K.-j. Kim, and I. Han, "Intelligent credit rating model for Korean companies using multiclass support vector machines," *Korean Management Review*, vol. 35, no. 5, pp. 1479–1496, 2006.

[2] H. Ahn and K.-j. Kim, "Corporate bond rating using various multiclass support vector machines," *Asia Pacific Journal of Information Systems*, vol. 19, no. 2, pp. 157–178, 2009.

[3] K.-j. Kim and H. Ahn, "A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach," *Computers & Operations Research*, vol. 39, no. 8, pp. 1800–1811, 2012.

[4] H. Ahn, "Optimization of multiclass support vector machine using genetic algorithm: Application to the prediction of corporate credit rating," *Information Systems Review*, vol. 16, no. 3, pp. 161–177, 2014.

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

[6] A. C. Lorena and A. P. L. F. de Carvalho, "Evolutionary tuning of SVM parameter values in multiclass problems," *Neurocomputing*, vol. 71, nos. 16–18, pp. 3326–3334, 2008.

[7] M.-D. Shieh and C.-C. Yang, "Multiclass SVM-RFE for product from feature selection," *Expert Systems with Applications*, vol. 35, nos. 1–2, pp. 531–541, 2008.

[8] S. Chatterjee, "Vision-based rock-type classification of limestone using multi-class support vector machine," *Neurocomputing*, vol. 39, no. 1, pp. 14–27, 2013.

[9] J. Chen, C. Zhang, X. Xue, and C.-L. Liu, "Fast instance selection for speeding up support vector machines," *Knowledge-Based Systems*, vol. 45, pp. 1–7, 2013.

[10] H. Ahn and K.-j. Kim, "Global optimization of case-based reasoning for breast cytology diagnosis," *Expert Systems with Applications*, vol. 36, no. 1, pp. 724–734, 2009.

[11] H. Ahn and K.-j. Kim, "Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach," *Applied Soft Computing*, vol. 9, no. 2, pp. 599–607, 2009.

[12] H. Ahn, K. Lee, and K.-j. Kim, "Global optimization of support vector machines using genetic algorithms for bankruptcy prediction," *Lecture Notes in Computer Science*, vol. 4234, pp. 420–429, 2006.

[13] G. E. Pinches and K. A. Mingo, "A multivariate analysis of industrial bond ratings," *Journal of Finance*, vol. 28, no. 1, 1973, pp. 1–18.

[14] A. Belkaoui, "Industrial bond ratings: A new look," *Financial Management*, vol. 9, no. 3, pp. 44–51, 1980.

[15] L. H. Ederington, "Classification models and bond ratings," *The Financial Review*, vol. 20, no. 4, pp. 237–262, 1985.

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

[17] J. Moody and J. Utans, "Architecture selection strategies for neural networks application to corporate bond rating," in *Neural Networks in the Capital Markets*, A. Refenes, Ed., Chichester: Wiley, 1995, pp. 277–300.

[18] J. J. Maher and T. K. Sen, "Predicting bond ratings using neural networks: a comparison with logistic regression," *Intelligent Systems in Accounting, Finance and Management*, vol. 6, no. 1, pp. 59–72, 1997.

[19] Y. S. Kwon, I. Han, and K. C. Lee, "Ordinal pairwise partitioning (OPP) approach to neural networks training in bond rating," *Intelligent Systems in Accounting, Finance and Management*, vol. 6, no. 1, pp. 23–40, 1997.

[20] R. Chaveesuk, C. Srivaree-Ratana, and A. E. Smith, "Alternative neural network approaches to corporate bond rating," *Journal of Engineering Valuation and Cost Analysis*, vol. 2, no. 2, pp. 117–131, 1999.
[21] K. S. Shin and I. Han, "A case-based approach using inductive indexing for corporate bond rating," *Decision Support Systems*, vol. 32, no. 1, pp. 41–52, 2001.
[22] L. Cao, L. K. Guan, and Z. Jingqing, "Bond rating using support vector machine," *Intelligent Data Analysis*, vol. 10, no. 3, pp. 285–296, 2006.
[23] W.-H. Chen and Shih J.-Y., "A study of Taiwan's issuer credit rating systems using support vector machines," *Expert Systems with Applications*, vol. 30, no. 3, pp. 427–435, 2006.
[24] Y.-C. Lee, "Application of support vector machines to corporate credit rating prediction," *Expert Systems with Applications*, vol. 33, no. 1, pp. 67–74, 2007.
[25] V. Vapnik, *The Nature of Statistical Learning Theory*. New York, NY: Springer–Verlag, 1995.
[26] U. Kreβel, "Pairwise classification and support vector machines," in *Advances in Kernal Methods: Support Vector Learning*, B. Scholkopf, C. Burges, and A. J. Smola, Eds., Cambridge, MA: MIT Press, 1999, pp. 255–268.
[27] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin, and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, no. 5, pp. 631–643, 2005.
[28] J. C. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin DAG's for multiclass classification," in *Advances in Neural Information Processing Systems 12*, S. A. Solla, T. K. Leen, and K.-R. Muller, Eds., Cambridge, MA: MIT Press, 2000, pp. 547–553.
[29] J. Weston and C. Watkins, "Support vector machines for multiclass pattern recognition," in *Proc. of the Seventh European Symposium on Artificial Neural Networks*, 1999, pp. 219–224.
[30] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
[31] T. Hong and J. Park, "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*, vol. 21, no. 2, pp. 43–58, 2011.
[32] F. E. H. Tay and L. Cao, "Application of support vector machines in financial time series forecasting," *Omega*, vol. 29, pp. 309-317, 2001.