A Hybrid Support Vector Machine Ensemble Model for Credit Scoring

Ahmad Ghodselahi
Department of Information Technology Management
Tarbiat Modares University
Tehran, Iran

ABSTRACT
Credit risk is the most challenging risk to which financial institution are exposed. Credit scoring is the main analytical technique for credit risk assessment. In this paper a hybrid model for credit scoring is designed which applies ensemble learning for credit granting decisions. The hybrid model combines clustering and classification techniques. Ten Support Vector Machine (SVM) classifiers are utilized as the members of ensemble model. Since even a small improvement in credit scoring accuracy causes significant loss reduction, then the application of ensemble in hybrid model leads to better performance of classification. A real dataset is used to test the model performance. The test results shows that proposed hybrid SVM ensemble has better classification accuracy when compared with other methods.

Keywords
Keywords-component: Ensemble, SVM, Credit scoring, Hybrid model

1. INTRODUCTION
Banking is special industry that deals with capital and risk for making profit. The bank success is directly pertaining to its capability of controlling and managing related risks. Banks are exposed to different kinds of risk, but the most challenging risk which can cause a bank to full failure is credit risk. The recent world's financial crisis has aroused remarkable consideration of financial institutions and banks on credit risk. Credit risk is an important and widely studied topic in the bank industry lending decisions and profitability. For all banks, credit remains the single largest risk being difficult to compensate. Credit risk is the loss of bank's profit, since the customer does not adhere to his or her loan refund commitment [1]. Usually, the generic approach of credit risk assessment is to apply some classification techniques on similar data of previous customers, both faithful and irresponsible customers, in order to find a relation between the characteristic and potential failures [2]. Credit scoring has become one of the main analytical ways for financial institutions to assess credit risk. The purpose of credit scoring is to classify the applicants into two groups: applicants with good credit and applicants with bad credit. Applicants with good credit have great possibility to repay financial obligation while, applicants with bad credit have high possibility of defaulting. The accuracy of credit scoring is critical for financial institution's profitability. Even 1% of improvement on the accuracy of credit scoring of applicants, will decrease a great loss for financial institutions. Usually credit score is a value that reveals the credit of the customer based on quantitative analysis of customer's credit history and characteristics [3]. The credit scoring model identifies financial variables that have statistical explanatory power in differentiating bad customers from good ones. The benefits obtained by developing a reliable credit scoring system are [4]:

- reducing the cost of credit analysis
- enabling faster decision
- insuring credit collections and diminish possible risk

Many organizations in the credit industry are developing new models to support the credit decisions. The objective of these new credit scoring models is increasing accuracy, which means more credit worthy applicants are granted credit, and consequently, increasing profits. The first credit scoring model was designed by Altman [5]. The credit scoring models can be divided into two categories: traditional models and novel models. The most common and utilized traditional models are Linear Discriminant Analysis (LDA) and Logistic Regression (LR) [6, 7, 8]. The weakness of the LDA is the assumption of linear relationship between variables, which is usually nonlinear and the sensitivity to the deviation from the multivariate normality assumption. The LR is predicting dichotomous outcomes and linear relationship between variables in the exponent of the logistic function, but does not require the multivariate normality assumption. Because of the deficiency, the linear relationship between variables, both LDA and LR are stated to have lack of accuracy [9]. Advances in information technology have lowered the cost of acquiring, managing and analyzing data, in an effort to build more robust and strong financial systems [10]. Recently, new approaches were applied for developing robust credit scoring systems. Recent studies have revealed that emerging artificial intelligent techniques, such as SVM, genetic algorithm and artificial neural networks are advantageous to statistical models and optimization technique for credit risk evaluation. Among the new techniques for credit scoring, SVM is one of those which generate prolific and promising results. The use of SVM in business application has been previously investigated by several works [11, 12, 13]. Baesens et al. [14] conducted a study for benchmarking of 17 different classification techniques on eight different real-life credit datasets. they used SVM and LS-SVM with linear and RBF kernels and adopted a grid search mechanism to tune the hyperparameters in their study. their experimental results indicated that SVM has the highest average ranking on performance. Schebesch and Stecking [15] used a standard SVM with linear and RBF kernel for applicant credit scoring and used a linear-kernel-based SVM to divide a set of labeled credit applicants into subsets of typical and critical patterns, which can be used for rejected applicants. In [16] SVMs were used for bankruptcy prediction and better accuracy was generated by SVM when compared to other methods. Gestel et al. [17] used LS-SVM for credit rating of banks, and compared the results with oridinary least squares, LR and multilayer perceptron (MLP). The results shows the LS-SVM classifier's accuracy is better than the other three methods. Min et al. [18] proposed methods for improving SVM performance in two aspects: feature subset selection and parameter optimization.

Although almost all classification methods can be used to assess credit risk, some hybrid approaches have shown higher correctness of predictability than any individual methods. In machine learning, the hybridization approach has been an active research area to improve classification or prediction performance over single learning approach. In general, it is based on combining two different machine learning techniques. For example, a hybrid classification model can be composed of one unsupervised learner to pre-process the training data and one supervised learner to learn the clustering result or vice versa. In [19], a hybrid mining approach was presented for credit scoring. Due to unrepresentative data, a two-stage approach was used which utilized self organizing map for clustering and neural networks to construct credit scoring model. Huang et al. [20] designed hybrid svm-based credit scoring models for assessing the credit scores of applicants. In [21] a hybrid credit scoring model was developed applying genetic programming and SVM. The accuracy of their hybrid model was better when compared with SVM, genetic programming, decisin tree, LR, and back propagation neural networks. Chen et al. [22] designed a SVM based hybrid model which mainly has three strategies. First using Cart, then using Mart and at last using grid search for model variable improvement. Tsai and Chen [23] present four hybrid credit scoring model and compare the performance of these hybrid models. Motivated by the hybrid model, integrating multiple classifier into aggregated output, ensemble learning, has been turned out to be an efficient method for achieving high classification performance. There is a growing interest that existing application of single classifier can be further improved by ensemble methods. The works of [3, 24, 25] have shown that ensemble methods have performed better than single classifier. Tsai and Wu [4] investigated the performance of a single classifier as the baseline classifier to compare with multiple classifiers and diversified multiple classifiers by NN based on three datasets. Zhou and Lai [26] developed multi-agent ensemble model for credit risk evaluation. Each agent is acted by a weighted LS-SVM. The results showed that the proposed model has better accuracy than other methods with which the proposed model was compared. Nanni and Lumini [25] compared the performance of ensemble of classifiers with single ones. The result showed applying ensemble method lead to better performance of classification. In [27], a comparative assessment of the performance of three popular ensemble methods- Bagging, Boosting, and Stacking- on credit scoring problem was conducted. The rest of this paper is organized as follows. In Section 2, an overview of ensemble learning is presented. In section 3, the details of experimental design are presented. Section 4 reports experimental results. Based on the observations and results of these experiments, section 5 draws conclusions.

2. OVERVIEW OF ENSEMBLE LEARNING

Ensemble learning is a machine learning paradigm where multiple learners are trained to solve the same problem. In contrast to ordinary machine learning approaches that try to learn one hypothesis from training data, ensemble methods try to construct a set of hypotheses and combine them to use [27]. This method is used to improve the performance and accuracy of classification task. The multiple classifier systems are based on the aggregation of a pool of classifiers such that their fusion achieves higher performance than the single classifiers. The key idea of most methods for building ensemble of classifiers is to modify the training dataset, builds classifiers on these *n* new training sets and then combines them into a final decision rule [25]. The rationale is that it may be more difficult to optimize the design of a single complex classifier than to optimize a combination of relatively simple classifiers. In addition, in ensemble models the error and deviation of one classifier is compensated by the other members of ensemble on classification task. The generalization ability of an ensemble is usually much stronger than that of a single learner, which makes ensemble method very attractive. In practice, to achieve a good ensemble, two necessary conditions should be satisfied: accuracy and diversity.

3. THE DESIGN OF THE METHODOLOGY

This section introduces the process of credit scoring model proposed by this study. The hybrid model employs two machine learning techniques, which is a combination of a clustering and a classification techniques. The process consists of Fuzzy C-Means clustering, normalization, building SVM classifier agents and finally, defining a method to combine the results generated by each agent. Each part of the hybrid credit scoring model is briefly described in following sub sections.

3.1. Fuzzy C-Means clustering

The first phase of the model is fuzzy clustering of dataset. This phase is as a pre-process for building SVM agents that generates homogeneous clusters with same features. This pre-process will lead to better training of SVM agents and as a result, better classification model is made and the probability of misclassification is reduced which is caused by inapt training data. Sometimes, even with a correct classification model, the ability of a model for predicting a new instance is limited. Such limitations are because of improper classification patterns which arise from training data. In addition, data uncertainty leads to more complex learning process of SVM agents. Therefore, the higher quality of training data causes the higher ability of SVM agents for correct classifications. The proposed model utilized FCM clustering to generate 10 clusters associated with their SVM agents. FCM is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 \quad , 1 \le m \le \infty \qquad (1)$$

where *m* is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster *j*, x_i is the *i*th of d-dimensional measured data, c_j is the d-dimensional center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center. Fuzzy portioning is carried out through an iterative optimization of the objective function (1), with the update of membership u_{ij} and the cluster c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad , \quad c_j = \frac{\sum_{i=1}^{N} u_{ij}^m . x_i}{\sum_{i=1}^{N} u_{ij}^m} \qquad (2)$$

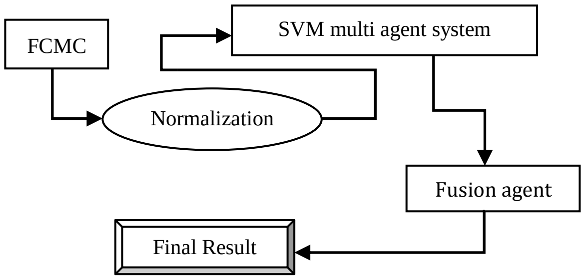

Fig 1: Proposed hybrid model

This iteration will stop when,

$$max_{ij}\left\{\left|u_{\mathrm{ij}}^{(\mathrm{k+1})} - \mathrm{u}_{\mathrm{ij}}^{(\mathrm{k})}\right|\right\} < \in, \quad (3)$$

Where ∈ is a termination criterion between 0 and 1, whereas k is the iteration steps. This procedure converges to a local minimum or a saddle point of J_m. The algorithm is composed of the following steps [28, 29]:

1. Initialize $U = [u_{ij}]\ matrix, U^{(0)}$
2. at k*-step: calculate the centers vectors* $C^{(k)} = [c_j]\ with\ U^{(k)}$
3. Update $U^{(k)}, U^{(k+1)}$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{\frac{2}{m-1}}}$$

4. If $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$ *then STOP; otherwise return to step 2*

3.2. Normalization
Data normalization should be performed in order to feed the SVM agents with data ranging in same interval for each input node. In credit assessment the numerical values representing the attributes of an applicant vary significantly in value and if a simple normalization process is applied to whole dataset, some useful information may be lost. At the normalization phase, the input data are separately normalized to values between 0 and 1. This is achieved by finding the maximum or highest value within each input attribute for all 1000 instances in dataset and dividing all the values within that same attribute by the obtained maximum value. This is a simple but efficient normalization.

3.3. SVM Agents
SVM technique is a classification technique that as an AI technique has proven its performance in many fields, such as text categorization, credit risk, and bankruptcy prediction. The strength of this technique lies with its capability to model nonlinearity and resulting in complex mathematical models. SVMs are used to find an optimal hyper-plane which maximizes the margin between itself and the nearest training examples in the new high-dimensional space and minimizes the expected generalization error. In this study, 10 SVMs are employed as ensemble members. The aim of the proposed model is to make full use of knowledge and intelligence of the members of group to make a rational decision over a pre-determined set of criteria.

3.4. Fusion Agents
Majority vote is the most common and used method for combining the group members' result in ensemble models. Despite the ability of these methods is good for combining, another method have been used which leaded to better classification accuracy than the aforementioned fusion methods. Every agent is assigned a weight, according to the sum of its members' membership degrees. The weight of agent 1 is the sum of Cluster 1 members' membership degrees.

4. EXPERIMENTAL ANALYSIS
In order to test the performance of hybrid model of this paper, the real world German dataset is used which is presented in follow.

4.1 Real World Dataset
The German dataset is available at UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)). It contains 1000 instances, with 700 cases were granted credit and 300 cases were refused. In these instances, each case is characterized by 20 decision attributes, 7 numerical and 13 categorical

4.2. Experimental Result
The evaluation criteria used to compare the tested methods are

- Total accuracy
- Type I accuracy
- Type II accuracy
- The area under the Receiver Operating Characteristic curve (AUC)

Different types of accuracy are calculated, based on the following equations:

$$Type\ I = \frac{Number\ of\ classified\ and\ also\ observed\ bad}{Number\ of\ observed\ bad}$$

$$Type\ I = \frac{Number\ of\ classified\ and\ also\ observed\ bad}{Number\ of\ observed\ bad}$$

$$Total\ accuracy = \frac{Number\ of\ correct\ classification}{Numbr\ of\ total\ evaluation}$$

In addition to calculating accuracy of each method, AUC of each method is calculated as well. In order to rank all models, the area under the receiver operating characteristic (ROC) graph is used as another performance measurement. The ROC graph is a useful technique for ranking models and visualizing their performance. Usually, ROC is a two-dimensional graph in which sensitivity is plotted on the Y-axis and 1-specificity is plotted on X-axis. Actually, the sensitivity is equal to type II accuracy and the specificity is equal to type I accuracy. To perform the model ranking task, a common method is to calculate the area under the ROC curve, abbreviated as AUC. Since the AUC is a portion of the area of the unit square, its value is always between 0 and 1. AUC can well describe the general behavior of the classification model because it is independent of any cutoff or misclassification cost used for obtaining a class label [30]. Generally, a model with a large AUC will have a good average performance. For each SVM agent, 4 types of kernel, namely RBF, polynomial, sigmoid and linear, are available. A kernel function can be interpreted as a kind of similarity measure between the input objects. Although some kernels are domain specific, there is in general no best choice. Since each kernel has some degree of variability in practice, there is nothing else for it but to experiment with different kernels. In this paper, four aforementioned kernel types were set in order to achieve best kernel for credit risk assessment model. For each agent the parameter C was set to 10 and gamma for RBF kernel was set to 0.1 and for other kernels, it was set to 1. Moreover, for each agent implementation with different kernel, the results of SVM agents were combined with both fusion methods which are majority vote and proposed method, membership degree. In this way the best fusion method will be defined based on the correct classification results. As the total accuracy of the hybrid model in each implementation with 4 types of kernels was shown in table 1, it is obvious the best total accuracy was generated when the SVM kernel is set to polynomial. Moreover, as it is showed in table 1, the polynomial and RBF kernels have generated highest total accuracy, respectively, when compared to other kernels. Linear kernel did not lead to good classification accuracy, for the relationship between the features is not linear. The comparison between two fusion methods' results showed that except in linear kernel that there is a slight difference between experimented fusion methods, in other kernel types, the membership degree fusion method resulted in better accuracy than majority vote; therefore, the new proposed ensemble member result combining method has better performance than the common combining method. According to the analysis of different structure for final model, the model was implemented with polynomial kernel and membership degree fusion method, was set for combining the results of ensemble members. In order to evaluate the performance of proposed model, other method for credit scoring is employed. The result of [27] for comparing common and popular methods of ensemble-bagging, boosting and stacking- with the result of model, is also presented. It is obvious that the proposed hybrid ensemble model has better performance. Its total accuracy is the best among mentioned methods and it surpasses other common ensemble methods. As it is presented in the table 2, proposed ensemble SVM can reach to better performance rather than the bagging SVM and boosting SVM. The ensemble of SVM has better performance when is compared to individual SVM and application of ensemble model improved the performance of SVM classifiers. When it comes to performance, the AUC is a good measurement criterion for classifiers. According to the AUC results, the ensemble of SVM has the best performance which makes it the first classifier in credit scoring models performance ranking.

Table 1. SVM kernels' total accuracy

SVM Kernel	Fusion method	
	Majority vote	Membership degree
	Total accuracy (%)	Total accuracy (%)
RBF	77.5	78.93
Polynomial	79.64	81.42
Sigmoid	49.64	56.42
Linear	71.78	71.07

Table 2. Result of credit scoring models (German dataset)

Method	Accuracy (%)			AUC
	Type I	Type II	Total	
DA	67.49	64.91	65.91	66.2
LR	48.68	84.67	71	66.675
Decision tree	49.93	78.55	70.35	64.24
RBFN individual	39.47	86.29	68.5	62.88
SVM individual	27.63	97.58	71	62.605
MLP individual	55.26	85.48	74	70.37
Bagging DT	48.28	86.34	74.92	67.31
Bagging NN	48.40	87.20	75.56	67.8
Bagging SVM	43.42	89.86	75.93	66.64
Boosting DT	40.89	82.96	72.77	61.925
Boosting NN	49.20	83.63	73.3	66.415
Boosting SVM	45.62	89.44	76.3	67.53
Stacking	45	89.24	75.97	67.12
Ensemble SVM (membership degree)	66.66	88.08	81.42	77.37

5. CONCLUSION

In this paper a hybrid model was developed which applies ensemble learning method to improve the performance of classification in the field of credit risk assessment. 10 SVM agents were utilized as the member of ensemble. For combining the result of each SVM agent, a new method was employed to fuse the result in ensemble model that outperform the common method used in credit scoring ensemble method, recently. In term of the model result, SVMs are apt classifier which can result in accurate classification; moreover, the ensemble of SVMs has better performance than the single SVM. The credit scoring results measured in this research support the hypothesis that ensemble of SVM can be used in credit scoring applications to improve the overall accuracy from a fraction of a percent to several percent. In an overall view, the proposed hybrid ensemble model that uses membership degree method to combine the result of ensemble members has the best accuracy and performance.

6. REFERENCES

[1]. Ya-qiong, P. 2007. A study on evaluation of consumer credit's risks of commercial banks. In Proceeding of International Conference on Wireless Communications (WiCom 2007). pp. 4531-4534.

[2]. Yu, L., Wang, S. A. and Lai, K. K. 2008. Credit risk assessment with a multistage neural network ensemble learning approach. Expert systems with applications. vol. 34. pp. 1434-1444.

[3]. Yu, L., Wang, S. and Lai, K. K. 2009. An intelligent-agent-based fuzzy group decision making model for financial multicriteria decision support: the case of credit scoring. European journal of operational research. vol. 195. pp. 942-959.

[4]. Tsai, C.-f. and Wu, J.-w. 2008. Using neural network ensembles for bankruptcy prediction and credit scoring. Expert systems with applications. vol. 34. pp. 2639-2649.

[5]. Altman, I. E. 1968. Financial ratios, discriminant analysis and the prediction of corporate banlruptcy. The journal of finance. vol. 23. pp. 589-611.

[6]. Baesens, B., Setiono, R., Mues, C. and Vanthienen, J. 2003. Using neural network rule extraction and decision tables for credit-risk evaluation. Management science. vol. 49. pp. 312-329.

[7]. Lee, T. S., and Chen, I. F. 2005. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. Expert sysytems with application. vol. 28. pp. 743-752.

[8]. West, D. 2000. Neural network credit scoring models. Computers and operations research. vol. 27. pp. 1131-1152.

[9]. Sustersic, M., Mramor, D. and Zupan, J. 2009. Consumer credit scoring models with limited data. Expert system with application. vol. 36. pp. 4736-4744.

[10]. Angelini, E., Tollo, G. D. and Roil, A. 2008. A neural network approach for credit risk evaluation. The quarterly review of economics and finance. vol. 48. pp. 733-755.

[11]. Wang, Y.-q. 2008. Building credit scoring systens based on support-based support vector machine. In proceeding of the Fourth international conference on natural computation. (ICNC 2008). pp. 323-327.

[12]. Huang, L. and Dai, Y. 2005. A support vector machine approach for prediction of T cell epitopes. In proceeding of the third asia-pacific bioinformatics conference. (APBC 2005). pp. 312-328.

[13]. Min, J. H. and Lee, Y.-C. 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. Expert systems with applications. vol. 28. pp. 603-614.

[14]. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J. and Vanthienen, J. 2003. Benchmarking state-of-art classification algorithm for credit scoring. Journal of operational research society. vol. 54. pp. 627-635.

[15]. Schebesch, K. B. and Stecking, R. 2005. Support vector machine for classifying and describing credit applicants: Detecting typical and critical regions. Journal of the operational research society. vol. 56. pp. 1082-1088.

[16]. Shin, K. S., Lee, T. S. and Kim, H. 2005. An application of support vector machines in bankruptcy prediction model. Exper systems with pplications. vol. 28. pp. 127-135.

[17]. Gestel, T. V., Baesens, B., Suykens, J. A., Van den Poel, D., Baestaens, D.-E. and Willekens, B. 2006. Bayesian kernel based classification for financial distress detection. European journal of operational research. vol. 172. pp. 979-1003.

[18]. Min, S. H., Lee, J. and Han, I. 2006. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. Expert systems with applications. vol. 31. pp. 652-660.

[19]. Hsieh, N. C. 2005. Hybrid mining approach in design of credit scoring model. Expert systems with applications. vol. 28. pp. 655-665.

[20]. Huang, C. L., Chen, M. C. and Wang, C. J. 2007. Credit scoring with a data mining approach based on support vector machines. Expert systems with applications. vol. 33. pp. 847-856.

[21]. Zhang, D., Hifi, M., Chen, Q. and Ye, W. 2008. A hybrid credit scoring model based on genetic programming and support vector machines. In proceeding of the fourth international conference on natural computation. (ICNC 2008). pp. 8-12.

[22]. Chen, W., Ma, C. and Ma, L. 2009. Mining the customer credit using hybrid support vector machine technique. Expert systems with applications. vol. 36. pp. 7611-7616.

[23]. Tsai, C.-F. and Chen, M. L. 2010. Credit rating by hybrid machine learning techniques. Applied soft computing. vol. 10. pp. 374-380.

[24]. Amin, M. F., Islam, M. M. and Murase, K. 2009. Ensemble of single-layered complex-valued neural networks for classification tasks. Neurocomputing. vol. 72. pp. 2227-2234.

[25]. Nanni, L. and Lumini, A. 2009. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring. Expert systems with applications. vol. 36. pp. 3028-3033.

[26]. Zhou, L. and Lai, K. K. 2009. Multi-agent ensemble models based on weighted least square SVM for credit risk assessment. In proceeding of the Global congress on intelligent systems. (GCIS 2009). pp. 559-563.

[27]. Wang, G., Hao, J., Ma, J. and Jiang, H. 2011. A comparative assessment of ensemble learning for credit scoring. Expert systems with applications. vol. 38. pp. 223-230.

[28]. Bezdek, J. C. Pattern recognition with fuzzy objective function algorithm. Newyork: Plenum press.

[29]. Dunn, J. C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of cybernetics and systems. vol. 3. pp. 32-57.

[30]. Fawcett, T. 2004. ROC Graphs: Notes and practical considerations for researchers. HP laboratories Palo Alto: Intelligent enterprise technologies laboratory.