Building Credit Scoring Systems Based on Support-based Support Vector Machine Ensemble

Yong-qiao Wang
College of Finance, Zhejiang Gongshang University, Hangzhou 310018, China
Wangyq@zjgsu.edu.cn

Abstract

This paper proposes a new strategy – support-based SVM ensemble for building credit scoring systems. Different from the commonly used "one-member-one-vote" majority-ruled ensembles, our proposed new framework aggregates degrees of support, or confidence levels, of several SVM classifiers to generate the final classification results that represent the consensus of the SVM. Decision values of a member SVM classifier are a good measurement of its support to positive or negative classification of an unlabeled sample. Two publicly available credit dataset have been used to test the usefulness and predicting power of the new approach. Results of both tests indicated clearly that the new approach outperformed the other three commonly used approaches: single, single best, and majority-rule ensemble.

1. Introduction

At the beginning of 2005, the total outstanding consumer credit reached $2.127 trillion in U.S. According to the Administrative Office of the U.S. Courts, in 2003 U.S. bankruptcy filings set a new record, totaling 1,660,245, which included 1,625,208 non-business and 35,037 business bankruptcy filings. Also, 2,062,000 people filed for bankruptcy in the year ended on December 31, 2004. Given such phenomenal exposure, evaluation of applications for credit lines, as well as credit monitoring after granting of credit lines and loans have received intense attention from the risk management community.

Many models and methodologies have been applied to support credit scoring, including statistical methods, optimization methods, nearest-neighbor approach, classification tree method, genetic algorithms, neural networks. A good recent survey on credit scoring and behavioral scoring can be found at Thomas (2002).

Support Vector Machine (SVM) has been developed and has proved to be useful in supporting credit scoring and other tasks that need "classification", include Huang et al. (2004) and Wang et al. (2005) etc. Huang et al. (2004) uses two datasets from Taiwanese financial institutions and United States' commercial banks as an experimental test bed to compare the performance of SVM with back propagation neural networks. Their results showed that SVMs achieved accuracy comparable with that of back propagation neural networks. Wang et al. (2005) proposes a new fuzzy support vector machine to evaluate credit risk which treats every training sample as both good and bad creditor, but with different memberships.

This paper proposes a new SVM ensemble method that aggregates the SVM classifiers of individual members by their degrees of "support" or "confidence" - which is determined by the decision value of the SVM classifier and makes it different from the "one-member-one-vote" system that used to be commonly used in SVM ensembles.

The rest of the paper is organized as follows. Section 2 is a simple introduction of SVM and SVM ensembles. Section 3 presents the new ensemble strategy in detail. Section 4 presents experimental results on four publicly available credit datasets. We conclude the paper in section 5.

2. SVM and SVM ensemble

2.1 SVM

The SVM algorithm operates by mapping the input space into a possibly higher-dimensional feature space and locating a plane that separates the positive examples from the negative ones as "far" as possible. Considering the following training samples: $\{x_k, y_k\}$, $k = 1, 2, \cdots N$, where $x_k \in R^d$ is the k_{th} input pattern, d is dimension of the input space and y_k is its corresponding observed result, 1 or -1. In credit

978-0-7695-3304-9/08 $25.00 © 2008 IEEE
DOI 10.1109/ICNC.2008.763

scoring, x_k denotes the attributes of applicants or creditors. y_k is the observed result of whether the debt is repaid timely or whether her application is approved. If the k_{th} customer's credit application is rejected, he defaults on his debt after being approved, then $y_k = -1$, else $y_k = 1$.

Assume we map the input points into a high dimensional space and solve the linear classification in there. The above training problem can be transformed into the following quadratic programming problem:

$$\max_{\alpha} J(\alpha) = \sum_{k=1}^{N} \alpha_k - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$$
$$\text{subject to}: \sum_{k=1}^{N} \alpha_k y_k = 0 \tag{1}$$
$$0 \le \alpha_k \le C \quad for \quad k = 1, \ldots, N$$

where C is a positive constant and $K(\bullet,\bullet)$ is a kernel function. The choices of kernel include linear, polynomial and RBF.

After solving (1) and substituting the optimal weight w into the original classification problem, we obtain the following classifier:

$$F(x) = sign\left(\sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b\right), \tag{2}$$

and the decision value function:

$$f(x) = w^T \varphi(x) + b = \sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b \quad . \tag{3}$$

2.2 SVM ensemble

Hansen and Salamon (1990) provided a simple explanation for why combinations of classifiers can outperform single best classifiers. If classification results of each classifier are independent and accuracy reaches p, then the accuracy of classification of a M-members ensemble is:

$$\sum_{m>M/2}^{M} \binom{M}{m} p^m (1-p)^{M-m} \tag{4}$$

if the fusion uses the majority rule, or the plurality rule. If p is larger than 0.5, which is a natural assumption for binary-classification problems, (4) is monotonically increasing with M. Many experimental results also show that classification results of ensembles are often superior to those of the single best classifier (Breiman, 1996 etc.).

Though theoretical analysis shows that the SVM has a good generalization ability, some obstacles still exist that may retard its application. The first obstacle is the approximate algorithm in solving large-scale quadratic programming in training step (Kim et al., 2003). The SVMs achieve good generalization ability only if global optimal solutions could be obtained. Because of the complexity surrounding the quadratic programming in training step, it usually has to resort to approximate algorithms. Such approximations may decrease the performance of SVMs. The second obstacle is the parameters selection. How to determine the regularization parameter C and the parameters of kernel functions? No fast and systematic approach has found wide acceptance. Thirdly, the complexity of input space may confound its discriminating power. Usually the characters of credit applicants include continuous, binary-class, and multi-class variables. Due to the above obstacles, it can not be guaranteed that a single classifier can achieve the best performance.

Most of the academic research on SVM ensembles has used majority, or plurality rule. Let $N_j = \#\{k \mid F_m(x) = k\}$, j=1 or -1, that is, the number of members' SVMs that assign sample x class label j. According to majority rule, the final decision of the SVM ensemble is:

$$F_e(x) = \arg\max_j N_j = sign\left(\sum_{m=1}^{M} F_m(x)\right), \tag{5}$$

where $F_e(\bullet)$ is the final classification result of the SVM ensemble and $F_m(\bullet)$ is the classification result of the i_{th} SVM classifier. When the ensemble consists of even base member classifiers and the number of member classifiers that are for positive classification is equal to the number of member classifiers that are for negative classification, the ensemble chooses randomly between positive and negative classification.

3. Support-based SVM ensembles

In SVM the decision value $f_m(x)$ is a good representation of the support of the member SVM classifier. The larger the $f_m(x)$, the larger is the degree by which the member m is for positive classification and is against negative classification, and vice verse. If a sample is on the side of positive class in the higher-dimensional feature space, the farther it is from the separating hyperplane, the more confident the member is about the hypothesis that sample *x* is from positive class, and vice verse. If sample *x* falls into the midst of the two boundaries and is very close to the separating hyperplane, the member is believed to be neutral about the classification of sample *x*. The specific algorithm of the new SVM ensemble is as follows. The architecture of support-based SVM ensembles is shown in Figure 1.

Algorithm of support-based SVM ensemble

Given: training dataset TR and sample x that needs to be labeled.

STEP 1. Bootstrapping M training datasets, TR_1, TR_2, , TR_M from TR.

STEP 2. Training M SVMs with datasets, TR_1, TR_2, , TR_M by QP (1) and obtaining M member SVM classifiers.

STEP 3. Use decision value function $f(x)$ to obtain the M SVM members' decision values of sample x_k, $f_m(x_k)$, $f_m(x_k)$,…, $f_m(x_k)$.

STEP 4. Map decision values of member SVMs to degrees of support for positive classification from the member SVMs $s_1^+(x_k)$, $s_2^+(x_k)$,…, $s_M^+(x_k)$ and degrees of support for negative classification $s_1^-(x_k)$, $s_2^-(x_k)$,…, $s_M^-(x_k)$. By this, the degrees of support lie in the unit interval [0,1].

Linear:

$$s_m^+(x_k) = \frac{f_m(x_k) - \min_{k=1,\dots,K} f_m(x_k)}{\max_{k=1,\dots,K} f_m(x_k) - \min_{k=1,\dots,K} f_m(x_k)} \tag{6}$$

Bridge:

$$s_m^+(x_k) = \begin{cases} 1 & s_k > \overline{s}_m \\ \dfrac{f_m(x_k) - \underline{s}_m}{\overline{s}_m - \underline{s}_m} & \underline{s}_m < f_m(x_k) \le \overline{s}_m \\ 0 & s_k \le \underline{s}_m \end{cases} \tag{7}$$

Logistic:

$$s_m^+(x_k) = \frac{e^{a_m * f_m(x_k) + b_m}}{e^{a_m * f_m(x_k) + b_m} + 1} \tag{8}$$

Probit:

$$s_m^+(x_k) = \Phi\left(\frac{f_m(x_k) - \mu_m}{\sigma_m}\right) \tag{9}$$

If the degree of support for positive classification is $s_m^+(x_k)$, the degree of support for negative classification is $s_m^-(x_k) = 1 - s_m^+(x_k)$.

STEP 5. Aggregate the degrees of support to obtain the final classification decision of sample x_k by (10)-(13).

Maximum rule:

$$F_e(x_k) = \begin{cases} 1 & \text{if } \max_{m=1,\dots,M} s_m^+(x_k) \ge \max_{m=1,\dots,M} s_m^-(x_k) \\ -1 & else \end{cases} \tag{10}$$

Minimum rule:

$$F_e(x_k) = \begin{cases} 1 & \text{if } \min_{m=1,\dots,M} s_m^+(x_k) \ge \min_{m=1,\dots,M} s_m^-(x_k) \\ -1 & else \end{cases} \tag{11}$$

Mean rule:

$$F_e(x_k) = \begin{cases} 1 & \text{if } \sum_{m=1,\dots,M} s_m^+(x_k) \ge \sum_{m=1,\dots,M} s_m^-(x_k) \\ -1 & else \end{cases} \tag{12}$$

Product rule:

$$F_e(x_k) = \begin{cases} 1 & \text{if } \prod_{m=1,\dots,M} s_m^+(x_k) \ge \prod_{m=1,\dots,M} s_m^-(x_k) \\ -1 & else \end{cases} \tag{13}$$

4. Experiments

To evaluate the efficacy of the proposed support-based SVM ensemble, we conduct tests with two publicly available credit dataset. The classification results of the proposed ensemble model are compared with three other models: single, single best and majority-rule ensemble. The single model trains a SVM classifier with the total training dataset and applies the classifier to predict the class of unknown samples. The single best model is constructed as follows: first, randomly generate several subsets of the total training dataset; second, use each subset to train a SVM classifier and evaluate its performance on the validation dataset, which is the collection of the total training dataset minus the respective subsets and, finally, choose the classifier which achieves the best performance as the final classifier. For single best model the member M means that we choose the final best performance from M candidate SVM classifiers. The new ensemble method uses 4 fusion functions, so finally in our test 7 models are compared.

To decrease the bias resulting from the choice of split between training set and test set, we randomly divide each dataset into a test set and the total training set and repeat training and evaluation in Q times. The final performance evaluation is based on the average results of the Q tests. In this paper we set Q as 20.

In this paper the classification performances are measured by Type1 accuracy, Type2 accuracy and overall accuracy, which are the percent of correctly classified good customers, percent of correctly classified bad customers and the percent of correctly classified customers in the total respectively.For statistical significance the paper uses paired t-test to compare the average total accuracy. Note t-test is just a heuristic method for comparison because of the fact that the forty rates (twenty pairs) are not independent.

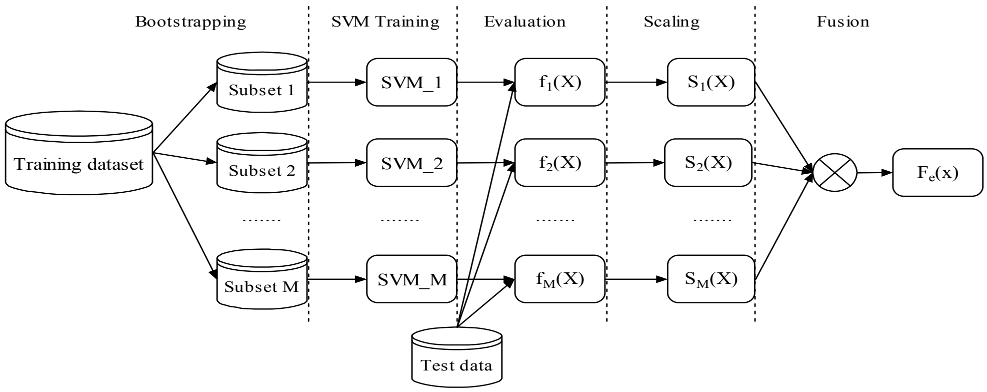

Figure 1. Architecture of support-based SVM ensembles

In this paper we use Probit function (9) to map the decision value of sample x obtained from a member SVM to degree of support for positive classification. Assume the percent of good creditors in the total training dataset TR_i is G% and the test dataset H_i has K samples. We let μ_m equal to the G% percentile of the decision values $f_m(x_1)$, $f_m(x_2)$,..., $f_m(x_K)$, which are obtained from the m_{th} member SVM classifier. And σ_m is the standard deviation of the decision values: $f_m(x_1)$, $f_m(x_2)$,..., $f_m(x_K)$.

4.1 Australian dataset

Australian Dataset concerns credit card applications and is available at http:// www.ics.uci.edu/ ~mlearn/ databases/ statlog/ australian. It consists of 307 and 383 samples of each class. Each sample is characterized by 14 attributes, 6 numerical and 8 categorical. In each SVM member we set the regularization parameter C at 0.5 and kernel as RBF with gamma that is randomly drawn from the interval [0.001, 0.01] over uniform distribution.

As shown in table 1, two models - support-based SVM ensemble with maximum fusion rule and support-based SVM ensemble with min rule - achieved the best performances (86.52%) among 7 models. The overall accuracies of the single model, the single best model, and majority-rule ensemble are 85.98%, 86.27% and 86.20%, respectively. By two-tail paired t-test, both the two best support-based SVM ensemble models achieved better performance than all other models at 10 percent significance level.

Figure 2 shows that, under most numbers of members, the supported SVM ensemble is superior to the single model, the single best model, and majority-rule ensemble model.

Table 1. Australian dataset performances

Model	Rule	Type1	Type2	Overall	Std
Single		84.47%	87.58%	85.98%	0.54%
Single best		84.81%	87.83%	86.27%	0.54%
Majority-rule ensemble		84.30%	88.08%	86.20%	0.00%
Support-based ensemble	Min	**85.52%**	**87.69%**	**86.52%**	2.72%
	Max	**85.52%**	**87.69%**	**86.52%**	2.72%
	Mean	85.13%	87.69%	86.34%	2.72%
	Product	85.13%	87.69%	86.34%	2.72%

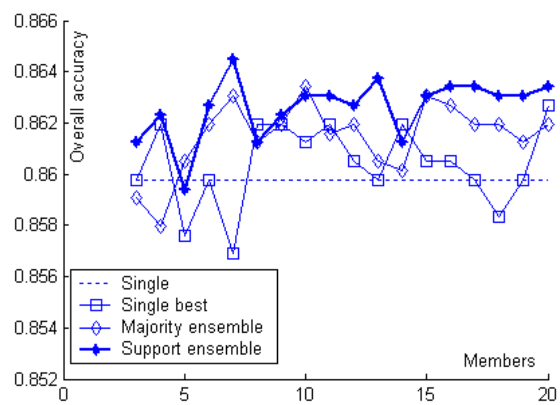

Figure 2. Australian dataset performances

4.2 German dataset

Professor Dr. Hans Hofmann of the University of Hamburg provided the German credit dataset that is available at http://www.ics.uci.edu/ ~mlearn/ databases/ statlog/ german/. It consists of 700 credit application samples that were accepted and 300 samples that were rejected. Among 20 variables 7 are numerical and 13 are categorical.

In each SVM member we set the regularization parameter C at 5 and kernel RBF with gamma that randomly drew from the interval [0.01, 0.05] over uniform distribution.

Table 2 shows the empirical results on German dataset. Two models - support-based SVM ensemble with mean fusion rule and support-based SVM ensemble with product rule - achieved the best performances (76.35%) among the eight models that were tested. The overall accuracies of the single model, the single best model, and majority-rule ensemble are 74.55%, 73.45% and 76.03%, respectively. By two-tail paired t-test, both the two best support-based SVM ensemble models achieved better performances than the single model and single best model at 10 percent significance level. The average performance of the best model is better than the majority-rule ensemble model, but without statistical significance at 10 percent level.

Table 2. German dataset performances

Model	Rule	Type1	Type2	Overall	Std
Single		80.64%	59.42%	74.55%	0.54%
Single best		79.87%	57.49%	73.45%	0.54%
Majority-rule ensemble		81.94%	61.35%	76.03%	0.00%
	Min	81.45%	59.55%	75.15%	2.72%
	Max	81.45%	59.55%	75.15%	2.72%
Support-based	Mean	**82.89%**	**60.17%**	**76.35%**	2.72%
ensemble	Product	**82.93%**	**60.08%**	**76.35%**	2.72%

Figure 3 shows that the supported SVM ensemble model with mean fusion rule achieved the best performance under most numbers of members. It is always better than the single model and the single best model and it is also better than or the same as the majority-rule ensemble under any number of members, except under 13 members. As a whole, the supported SVM ensemble model performed the best and the majority-rule ensemble model the second.

5. Summary and conclusions

In this paper, we proposed a new SVM ensemble strategy for credit analysis. Different from commonly used "one-member-one-vote" or "majority-rule" ensemble, the new SVM ensemble aggregates the decision values from the SVM members, instead of their classification results directly. Two publicly available credit dataset have been used to test the usefulness and predicting power of the new ensemble. All of them show clearly that the new ensemble outperformed the other three models (single, single best and majority-rule ensemble) on average. Moreover the superiority seemed to be consistent when the ensemble consists of different numbers of SVM members. The results also show that we can't upgrade the performance of the new SVM ensemble by adding homogeneous classifiers into the ensemble. Furthermore the new SVM ensemble strategy has a great potential to find its applications in many other binary-class classification problems. Researchers can apply the strategy without further modifications.

Further research required on the support-based SVM ensemble strategy is vast. How about the efficiency of the new ensemble if other scaling functions are used in scaling the step? Can trainable algorithms, such as neural networks, SVM itself or fuzzy approach etc., be used in fusion step to improve its performance?

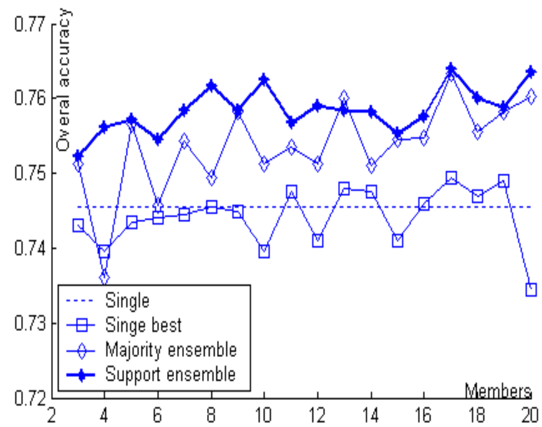

Figure 3. German dataset performances

References

[1] L. Breiman, "Bagging Predictors". Machine Learning, Vol. 24, pp. 123－140, 1996.

[2] L. Breiman, "Prediction Games and Arcing Algorithms", Neural Computation, Vol. 11, pp. 1493－1517, 1999.

[3] L. K. Hansen and P. Salamon, "Neural network ensemble", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 12, pp. 993-1001, 1990.

[4] Z.Huang, H. C. Chen, C. J. Hsu, W. H. Chen, S. S. Wu , "Credit Rating Analysis with Support Vector Machines and Neural Networks: a Market Comparative Study", Decision Support Systems, Vol. 37, pp. 543-558, 2004.

[5] H. C. Kim, S. N. Pang, H. M. Je, D. Kim and S. Y. Bang, "Constructing support vector machine ensemble", Pattern Recognition, Vol. 36, pp. 2757-2767, 2003.

[6] L. C. Thomas, "A Survey of Credit and Behavioural Scoring: Forecasting Financial Risk of Lending to Consumers", International Journal of Forecasting, Vol. 16, pp. 149-172, 2002.

[7] V. Vapnik, "Statistical learning theory", New York: Wiley. 1998.

[8] Y. Q. Wang, S. Y. Wang, K. K. Lai, "A New Fuzzy Support Vector Machine to Evaluate Credit Risk", IEEE Transactions on Fuzzy Systems, Vol. 13, pp. 820-831, 2005.