A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine

Gang Wang a,b,c,*, Jian Ma c

a School of Management, Hefei University of Technology, Hefei, Anhui, PR China
b Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education, Hefei, Anhui, PR China
c Department of Information Systems, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

Keywords:
Enterprise credit risk assessment
Ensemble learning
Bagging
Random subspace
SVM

ABSTRACT

Enterprise credit risk assessment has long been regarded as a critical topic and many statistical and intelligent methods have been explored for this issue. However, there are no consistent conclusions on which methods are better. Recent researches suggest combining multiple classifiers, i.e., ensemble learning, may have a better performance. In this paper, we propose a new hybrid ensemble approach, called RSB-SVM, which is based on two popular ensemble strategies, i.e., bagging and random subspace and uses Support Vector Machine (SVM) as base learner. As there are two different factors, i.e., bootstrap selection of instances and random selection of features, encouraging diversity in RSB-SVM, it would be advantageous to get better performance. The enterprise credit risk dataset, which includes 239 companies’ financial records and is collected by the Industrial and Commercial Bank of China, is selected to demonstrate the effectiveness and feasibility of proposed method. Experimental results reveal that RSB-SVM can be used as an alternative method for enterprise credit risk assessment.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

The enterprise credit risk assessment has long been regarded as important and widely studied issue in the academic and business community. In recent years, enterprise credit risk assessment has become one of the primary ways for financial institutions to assess credit risk, improve cash flow, reduce possible risks and make managerial decisions (Huang, Chen, Hsu, & Wu, 2004; Huang, Chen, & Wang, 2007; Wang, Hao, Ma, & Jiang, 2011). For the enterprise credit risk assessment, the accuracy is quite significant to financial institutions‘ profitability. For example, the accuracy of assessment increases only one percent may retrieve a great loss for financial institutions (Hand & Henley, 1997). Some statistical methods have been widely applied to build the enterprise credit risk assessment models, such as Linear Discriminant Analysis (LDA) (Karels & Prakash, 1987; Reichert, Cho, & Wagner, 1983), Logistic Regression Analysis (LRA) (Thomas, 2000; West, 2000), Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991). However, the problem with applying these statistical methods to enterprise credit risk assessment is that some assumptions, such as the multivariate normality assumptions for independent variables, are frequently violated in reality, which makes these methods theoretically invalid for finite samples (Huang et al., 2004).

In recent years, many studies have demonstrated that intelligent methods, such as Artificial Neural Network (ANN) (Desai, Crook, & Overstreet, 1996; West, 2000), Decision Tree (DT) (Hung & Chen, 2009; Makowski, 1985), Case Based Reasoning (CBR) (Shin & Han, 2001; Wheeler & Aitken, 2000) and Support Vector Machine (SVM) (Baesens et al., 2003; Huang et al., 2007; Schebesch & Stecking, 2005) can be used as alternative methods for enterprise credit risk assessment. In contrast with statistical methods, intelligent methods do not assume certain data distributions. These methods automatically extract knowledge from training data. According to previous studies, intelligent methods are superior to statistical methods in dealing with enterprise credit risk assessment problems, especially for nonlinear pattern classification (Huang et al., 2004). Among them, one of the most effective methods is SVM and has been successfully applied into enterprise credit risk assessment. However, the practical SVM has been implemented based on the approximation algorithm to reduce the cost of time and space. So, the obtained classification performance is far from the theoretically expected (Baesens et al., 2003; Huang et al., 2007; Schebesch & Stecking, 2005). Baesens et al. (2003) applied SVM, along with other classifiers to several enterprise credit risk datasets. They reported that SVM performs well in comparison with other algorithms, but do not always give the best performance. Schebesch and Stecking (2005) applied SVM to a database of applicants for building enterprise credit risk assessment model. They concluded...
that SVM performs slightly better than LRA, but not significantly so. Huang et al. (2007) found SVM classifies enterprise credit applications no more accurately than ANN, DT and Genetic Algorithms (GA), and compared the relative importance of using features selected by GA and SVM along with ANN and genetic programming.

To overcome these limitations, recently, integrating multiple classifiers into an aggregated output, i.e., ensemble method, has been turned out to be an efficient strategy for achieving high classification performance, especially when the base classifiers have different structures that lead to independent prediction errors (Breiman, 1996; Dietterich, 2000; Opitz & Maclin, 1999; Schapire, 1999; Wolpert, 1992). Yu, Wang, and Lai (2008) proposed a multi-stage neural network ensemble learning model to evaluate credit risk. Experimental results revealed the proposed neural network ensemble learning model can provide a promising solution to credit risk assessment. Tsai and Wu (2008) investigated the performance of a single classifier as the base learner to compare with multiple classifiers and diversified multiple classifiers by using neural networks. By comparing with the single classifier as the benchmark in terms of average accuracy, the ensemble method performs better. Nanni and Lumini (2009) investigated the performance of several systems based on ensemble methods for enterprise credit risk assessment. The results showed that ensemble methods may be used for boosting the performance of “stand-alone” classifier. Hung and Chen (2009) proposed a selective ensemble model of three classifiers, i.e., DT, ANN and SVM for enterprise credit risk assessment. Based on the expected probabilities of credit risk, this ensemble method provides an approach which inherits advantages and avoids disadvantages of different classification methods.

Base on the above motivation, we propose a new hybrid ensemble approach, called RSB-SVM, which is based on two popular ensemble strategies, i.e., bagging and random subspace and use SVM as base learner for enterprise credit risk assessment. Both theoretical and experimental researches show that combining a set of accurate and diverse classifiers will lead to a powerful classification system (Breiman, 1996; Dietterich, 2000; Opitz & Maclin, 1999; Schapire, 1999). For the first condition, accuracy, we choose SVM as the base learner. And for the diversity, among the diverse ensemble methods that are available, bagging and random subspace are two more often used methods and have been found to be accurate, computationally feasible across various data domain. In addition, it has been observed that an important prerequisite for ensemble methods to reduce the test error is that it generates a diversity of ensemble members (Breiman, 1996; Dietterich, 2000; Opitz & Maclin, 1999). However, for bagging, the only factor encouraging diversity is the proportion of different objects in the training samples. Although the classifier techniques used in bagging are sensitive to small changes in data, the bootstrap sampling appears to lead to ensembles of low diversity compared to other ensemble methods, e.g., boosting. In order to encourage diversity, we can use random subspace strategy to select a subset of features as input. As a result, we introduce random subspace strategy into Bagging SVM and get RSB-SVM. As there are two different factors, i.e., bootstrap selection of instances and random selection of features, encouraging diversity in RSB-SVM, it would be advantageous to get better performance. For the testing and illustration purposes, the enterprise credit risk dataset, which includes 239 companies’ financial records from China and is collected by the Industrial and Commercial Bank of China, is selected to demonstrate the effectiveness and feasibility of proposed method. Experimental results reveal that RSB-SVM performs the best among eight methods, i.e., SVM, Bagging SVM, Random Subspace SVM, Boosting SVM, LRA, DT and ANN. And the non-linear kernel of SVM is more feasible than the linear kernel in the enterprise credit risk assessment practice. All these results illustrate that RSB-SVM can be used as an alternative method for enterprise credit risk assessment.

The remainder of the paper is organized as follows. In Section 2, the background of SVM, bagging and random subspace are presented. In Section 3, we propose a new hybrid ensemble approach, RSB-SVM, based on the bagging and the random subspace for enterprise credit risk assessment. In Section 4, we present the details of experiment design and report experimental results. Based on the observations and results of these experiments, Section 5 draws conclusions and future research directions.

2. Background

2.1. Support Vector Machine

As a relatively new class of machine learning techniques based on statistical learning theory (Cortes & Vapnik, 1995; Vapnik, 1995), SVM for enterprise credit risk assessment has obtained several state-of-art results in classification accuracy. In SVM, original input space is mapped into a high-dimensional dot product space called a feature space, and in the feature space the optimal hyperplane is determined to maximize the generalization ability of the classifier. The optimal hyperplane is found by exploiting the optimization theory, and respecting insights provide by the statistical learning theory. Fig. 1 shows an illustration of the idea of an optimal hyperplane for linearly separable patterns.

Given a set of training samples \( D = \{(x_1, y_1), \ldots, (x_N, y_N)\} \), where \( x_i \in \mathbb{R}^d \) is the vector space pattern, \( y_i \in \{-1, 1\} \) is the class label for a 2-class problem, SVM for classification attempts to find a classifier \( f(x) \), which minimizes the expected misclassification rate. A linear classifier \( f(x) \) is a hyperplane, and can be represented as \( f(x) = \text{sgn}(w^T x + b) \).

Finding the optimal classifier \( f(x) \) in SVM is equivalent to solving a convex quadratic optimization problem in (1):

\[
\begin{align*}
\min_{w, b} & \quad \frac{1}{2}||w||^2 + C \sum_{i=1}^{N} \xi_i \\
\text{subject to} & \quad y_i (w^T x_i + b) \geq 1 - \xi_i \quad (\xi_i \geq 0, i = 1, \ldots, N)
\end{align*}
\]

(1a)

where \( C \) is called the regularization parameter, and is used to balance the classifier’s complexity and classification accuracy on the training set \( D \). This quadratic problem is generally solved through its dual formulation. Simple replacing the involved vector inner-product with a non-linear kernel function converts linear SVM into a more flexible non-linear SVM, which is essence of the famous kernel trick. Any function satisfying Mercer’s condition can be used as the kernel function (Vapnik, 1995).

Some typical kernel function are:

- Linear: \( K(x_i, x_j) = x_i^T x_j \)
- Polynomial: \( K(x_i, x_j) = (x_i^T x_j + 1)^d \)

Fig. 1. A linear separable Support Vector Machine.
2.2. Bagging

Breiman’s bagging, short for bootstrap aggregating, is one of the earliest ensemble learning algorithms (Breiman, 1996). It is also one of the most intuitive and simplest to implement, with a surprisingly good performance. Classifiers composing an ensemble are usually called base learners. Diversity in bagging is obtained by using bootstrapped replicas of the training dataset: different training data subsets are randomly drawn—with replacement—from the entire training dataset. Each training data subset is used to train a different base learner of the same type.

The base learners’ combination strategy for bagging is majority vote. Simple as it is, this strategy can reduce variance when combined with the base learner generation strategies. The pseudo-code for the bagging algorithm is given in Fig. 2.

2.3. Random subspace

The random subspace method is an ensemble construction technique proposed by Ho (1998). In the random subspace, the training dataset is also modified as in bagging. However, this modification is performed in the feature space (rather than instance space). The pseudo-code for the random subspace algorithm is given in Fig. 3.

The random subspace may benefit from using both random subspaces for constructing the base learners and aggregating the base learners. When the dataset has many redundant or irrelevant features, one may obtain better base learners in random subspaces than in the original feature space (Ho, 1998). The combined decision of such base learners may be superior to a single classifier constructed on the original training dataset in the complete feature space.

3. Hybrid ensemble approach RSB-SVM for enterprise credit risk assessment

Assess enterprise credit risk is a hot topic in management science due to its importance for making correct business decisions. Both statistical and intelligent methods have been explored for this issue. Although there are no consistent conclusions on which methods are better, recent researches suggest combining multiple classifiers, i.e., ensemble method, may have a better performance.

Ensemble method is a machine learning paradigm where multiple classifiers are trained to solve the same problem. In contrast to ordinary machine learning approaches which try to learn one hypothesis from training data, ensemble methods try to construct a set of hypotheses and combine them to use (Zhou, 2009). Many ensemble methods are able to boost weak classifiers which are slightly better than random guess to strong classifiers which can make very accurate predictions. However, it is noteworthy that although most theoretical analyses work on weak classifiers, base learners used in practice are not necessarily weak since using not-so-weak base learners often results in better performance. Thus based on the above literature analysis, we use SVM as base learner for enterprise credit risk assessment in this study.

Great improvement in generalization performance has been observed from ensemble method in a wide range of numerical experiments and practical applications. In practice, to achieve a good ensemble, two necessary conditions should be satisfied: accuracy and diversity. For the first condition, accuracy, we could simply mean that the base learner should be more accurate than random guessing. In this study, we use SVM as the base learner which is satisfied with the above condition. For the second condition, diversity, we mean that each base learner has its own knowledge about the problem and has a different pattern of errors compared to

---

**Fig. 2.** The bagging algorithm.

**Fig. 3.** The random subspace algorithm.
other base learners. Focusing on diversity, there are different methods for construction of diverse base learners. For example, just as mentioned above, bagging perturbs the distribution of the training set by resampling. And random subspace perturbs the feature space to get diversity. For bagging, however, there is the only factor encouraging diversity. Although the base learner used in bagging are sensitive to small changes in data, the bootstrap sampling appears to lead to ensembles of low diversity compared to other ensemble methods, e.g., boosting. To enforce diversity, a version of bagging called Random Forest was proposed by Breiman (2001). The ensemble consists of decision trees built again on bootstrap samples. The difference lies in the construction of the decision tree. The feature to split a node is selected as the best feature among a set of $M$ randomly chosen features, where $M$ is a parameter of the algorithm. This small alteration appeared to be a winning heuristic in that diversity was introduced without much compromising the accuracy of the base learners.

Even SVM has been proposed to provide a good generalization performance, the classification result of the practically implemented SVM is often far from the theoretically expected level because their implementations are based on the approximated algorithms due to the high complexity of time and space. To improve the limited classification performance of SVM, followed Breiman’s improvement (Breiman, 2001), we propose a new hybrid ensemble approach, called RSB-SVM, which is based on two popular ensemble strategies, i.e., bagging and random subspace and uses Support Vector Machine (SVM) as base learner. RSB-SVM firstly divides the dataset into some subsets using bootstrap sampling with replacement method in bagging. Subsequently, the new subsets are selected by random subspace method in random subspace and uses Support Vector Machine (SVM) as base learner. RSB-SVM firstly divides the dataset into some subsets using bootstrap sampling with replacement method in bagging. Subsequently, the new subsets are selected by random subspace method in random subspace from the original feature set. Then, it trains the different SVMs using different new subsets. Lastly it uses majority vote method in bagging to aggregate the results. Through above procedures, we introduce random subspace strategy into Bagging SVM and get RSB-SVM. The whole framework of RSB-SVM is illustrated in Fig. 4.

As there are two different ensemble strategies, i.e., bootstrap selection of instances and random selection of features, encouraging diversity in RSB-SVM, it would be advantageous to get more accuracy than bagging and random subspace individually. The pseudo-code for the RSB-SVM algorithm is given in Fig. 5.

4. Experimental analyses

4.1. Experimental design

In order to evaluate the performance of hybrid ensemble approach, RSB-SVM, the enterprise credit risk dataset is derived from 239 companies (148 non-risk cases and 91 risk cases) that were granted loans form the Industrial and Commercial Bank of China, a premier bank of China, between the year of 2006 and 2007. The dataset include these enterprises’ detailed financial records and corresponding results of assessment. In this study, 18 financial variables were chosen as the criteria for the credit risk assessment, according to the experts from the credit bureau of the Industrial and Commercial Bank in Shanghai. These variables cover their financial structure, their ability of paying debt, the management’s

![Fig. 4. Framework of RSB-SVM for enterprise credit risk assessment.](image)

![Fig. 5. The RSB-SVM algorithm.](image)
ability and the operations profitability, as listed in Table 1. The dependent variable is the credit status of the enterprise: risk or non-risk.

The evaluation criteria of our experiments are adopted from the established standard measures in the fields of credit risk assessment. These measures include average accuracy, type I error and type II error. Each measure has its merits and limitations. In this study, we prefer to use a combination of these measures, rather than a single measure, to measure the performance of enterprise credit risk assessment models. The definition of these measures can be explained with respect to a confusion matrix as shown in Table 2.

Formally speaking, they are defined as follows:

Average accuracy \( \frac{TP + TN}{TP + FP + FN + TN} \) (4)

Type I error \( \frac{FN}{TP + FN} \) (5)

Type II error \( \frac{FP}{TN + FP} \) (6)

To minimize the influence of the variability of the training set, ten times 10-fold cross validation is performed. In detail, the enterprise credit risk dataset is partitioned into ten subsets with similar sizes and distributions. Then, the union of nine subsets is used as the training set while the remaining subset is used as the test set, which is repeated for ten times such that every subset has been used as the test set once. The average test result is regarded as the result of the 10-fold cross validation. The whole above process is repeated for 10 times with random partitions of the ten subsets, and the average results of these different partitions are recorded.

4.2. Experimental results

The experiments described in this section were performed on a PC with a 1.83 GHz Intel Core Duo CPU and 2 GB RAM, using Windows XP operating system. Data mining toolkit WEKA (Waikato Environment for Knowledge Analysis) version 3.6.0 is used for experiment. WEKA is an open source toolkit, and it consists of a collection of machine learning algorithms for solving data mining problems (Witten & Frank, 2005).

In this study, we compared RSB-SVM with other seven common used methods in enterprise credit risk assessment, e.g., LRA, DT, ANN, SVM, bagging SVM, random subspace SVM (RS SVM) and Boosting SVM. For implementation of LRA, DT, ANN and SVM, we chose Logistic module, J48 (WEKA’s own version of C4.5) module, MultilayerPerceptron module and SMO (WEKA’s own version of SVM) module in WEKA. And for implementation of ensemble learning, i.e., bagging SVM, RS SVM and Boosting SVM, we chose Bagging module, RandomSubSpace module and ADBoostM1 module. For implementation of RSB-SVM, we used WEKA Package, i.e., WEKAJAR and implement in Eclipse. For base learner SVM, we chose the linear kernel and polynomial kernel \((d = 2)\) for the experiments. Except when stated otherwise, all the default parameters in WEKA were used. Moreover, five random subspace rates for RS SVM and RSB-SVM are tested, where the value of random subspace rate is set to 0.5, 0.6, 0.7, 0.8 and 0.9, respectively.

Table 3 presents three performance indicators, i.e., average accuracy, type I error and type II error, of different methods. For RS SVM and RSB-SVM, we report the results when the random subspace rate was set to 0.9. In addition, the experimental results on the other random subspace rate will be presented later in this subsection.

It is evident from the Table 3 that RSB-SVM (polynomial) has the highest average accuracy of 78.98%. Closely following RSB-SVM (polynomial) is Bagging SVM (polynomial) with an average accuracy of 74.97% and Boosting SVM (polynomial) with 74.63%. Note that SVM (polynomial) get the best average accuracy among five base learners, i.e., LRA (71.69%), DT (69.06%), ANN (71.52%), SVM (linear) (68.02%) and SVM (polynomial) (73.84%). This result illustrates the availability of choosing SVM as base learner for RSB-SVM.

As SVM (polynomial) gets better average accuracy than SVM (linear) and other four ensemble SVM (polynomial) methods also get better average accuracy than ensemble SVM (linear) methods, it is illustrated that the enterprise credit risk dataset is not separated linearly. These also make SVM (linear) classify more than 95% instances into non-risk cases and get lower type I error (2.76%) and higher type II error (79.48%). Thus in order to separate the enterprise credit risk dataset non-linearly, the polynomial kernel is more feasible in practice.

Subsequently, Figs. 6 and 7 display the average accuracy curve, type I error curve and type II error curve for SVM and other four ensemble SVM methods when the random subspace rate varies.
from 0.5 to 0.9. When the number of random subspace rate is small, the performance of RSB-SVM may be worse than SVM, bagging SVM and boosting SVM. However, with the increasing of the random subspace rate, the performance of RSB-SVM is becoming better than SVM, bagging SVM and boosting SVM. These results further prove that the combining two ensemble strategies, i.e., bagging and random subspace, can enhance the performance of enterprise credit risk assessment. It is interesting that RS SVM get worse accuracy than other methods. The reason may be that the enterprise credit risk dataset has less redundant or relevant features. These results may also suggest that independently using RS SVM in enterprise credit risk assessment is not feasible.

Subsequently, in order to ensure that the assessment does not happen by chance, we tested the significance of these results by means of the paired t-test. The null hypothesis is “Model A’s mean of Average accuracy/Type I error/Type II error = Model B’s mean Average accuracy/Type I error/Type II error”. The alternative hypothesis is “Model A’s mean Average accuracy/Type I error/Type II error ≠ Model B’s mean Average accuracy/Type I error/Type II error”. The column ‘improvement’ gives the relative improvement in mean Average accuracy (Type I error or Type II error) that Model A gives over Model B. The results are summarized in Table 4. As shown in Table 4, the proposed RSB-SVM (polynomial) is significantly better than all other 12 methods.
5. Conclusions and future directions

Enterprise credit risk assessment has become a very important task as financial institutions have to decide whether to grant credit to enterprises who submit an application. And only one percent assessment accuracy increase would retrieve a great loss for financial institutions. In this paper, a hybrid ensemble approach, called RSB-SVM, is proposed for enterprise credit risk assessment. This approach works through integration of two popular ensemble strategies, i.e., bagging and random subspace. RSB-SVM outperforms bagging and random subspace in generating more diverse component SVM classifiers. Experiments based on the credit risk dataset, which collected by the Industrial and Commercial Bank of China, demonstrate that RSB-SVM gets the best performance among the eight methods, i.e., SVM, Bagging SVM, Random Subspace SVM, Boosting SVM, LRA, DT and ANN. And in practice, the non-linear kernel of SVM is more feasible than the linear kernel for enterprise credit risk assessment.

Several future research directions also emerge. Firstly, we only use one dataset to validate the proposed method in this study. In the further study, large datasets for experiments and applications, particularly with more exploration of credit risk data structures, should be collected to further validate the conclusions of this study. Secondly, other kernel functions of SVM, such as radial basis function and sigmoid, could be further tested in the next research. Thirdly, the experimental results have shown that combining different ensemble strategies can achieve better performance. Thus, more extensive combination of ensemble strategies can be investigated in the future research.

Acknowledgments

This work is partially supported by the National Natural Science Foundation of China (No. 71071045, 71101042), the National High-tech R&D Program of China (863 Program) (No. 2009AA043403), the China Postdoctoral Science Foundation (2011M501041) and the Doctoral Special Fund of Hefei University of Technology (No. 2010HGBZ0607).

References