Jrl Syst Sci & Complexity (2008) 21: 527–539

DESIGNING A HYBRID INTELLIGENT MINING SYSTEM FOR CREDIT RISK EVALUATION*

Lean YU · Shouyang WANG · Fenghua WEN · Kin Keung LAI · Shaoyi HE

Received: 8 July 2008 / Revised: 2 September 2008

Abstract In this study, a novel hybrid intelligent mining system integrating rough sets theory and support vector machines is developed to extract efficiently association rules from original information table for credit risk evaluation and analysis. In the proposed hybrid intelligent system, support vector machines are used as a tool to extract typical features and filter its noise, which are different from the previous studies where rough sets were only used as a preprocessor for support vector machines. Such an approach could reduce the information table and generate the final knowledge from the reduced information table by rough sets. Therefore, the proposed hybrid intelligent system overcomes the difficulty of extracting rules from a trained support vector machine classifier and possesses the robustness which is lacking for rough-set-based approaches. In addition, the effectiveness of the proposed hybrid intelligent system is illustrated with two real-world credit datasets.

Key words Credit risk evaluation, hybrid intelligent system, rough sets, support vector machine.

1 Introduction

One of the most significant threats to many businesses today is their counterparts' credit risk, regardless of the sizes and the nature of their operations. Therefore, credit risk evaluation becomes extremely important for profit-earning and sustainable development of a firm. Due to its importance, various models are used for credit risk evaluation tasks. First of all, many statistical models and optimization techniques, such as linear discriminant analysis[1−2], logit analysis[2−3], probit analysis[4], linear programming[5], integer programming[6], k-nearest neighbor (k-NN)[7], and classification tree[8] are widely applied to credit risk assessment and

Lean YU · Shouyang WANG
Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, *China.* Email: yulean@amss.ac.cn; sywang@amss.ac.cn.
Fenghua WEN
School of Economy and Management, Changsha University of Science & *Technology, Changsha* 410076, *China.* Email: mailtowfh@126.com.
Kin Keung LAI
Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong. Email: mskklai@cityu.edu.hk.
Shaoyi HE
Department of Information Systems and Operations Management, College of Business Administration, California State University San Marcos, San Marcos, CA 92096-0001, *USA.* Email: she@csusm.edu.
*This research was partially supported by the National Natural Science Foundation of China under Grant Nos. 70221001, 70701035, the Knowledge Innovation Program of the Chinese Academy of Sciences under Grant Nos. 3547600, 3046540, 3047540, the Key Research Institute of Philosophies and Social Sciences in Hunan Universities, and the National Natural Science Foundation of China/Research Grants Council (RGC) of Hong Kong Joint Research Scheme under Grant No. N_CityU110/07.

modeling tasks. Although these methods can be used to assess credit risk, the ability to discriminate good customers from bad ones is still worth improving further. Recent studies have revealed that emerging artificial intelligence (AI) techniques, such as artificial neural networks (ANN)[9−12], genetic algorithm (GA)[13−14], support vector machine (SVM)[15−18], and rough set theory[19−20] are advantageous to statistical models and optimization techniques for credit risk evaluation.

Although almost all classification methods can be used to evaluate credit risk, some hybrid and combined classifiers, which integrate two or more single classification methods, have shown higher correctness of predictability than any individual methods, especially in fields where the development of a powerful single classification system requires considerable efforts. Studies on hybrid and combined classifier are currently flourishing in credit risk assessment. Recent examples are neural discriminant technique[21], neuro-fuzzy[22−23], rough-neural[24], and fuzzy SVM[25]. Some comprehensive literature review on credit risk assessment and modeling can be found in two recent surveys[26−27].

Inspired by the previous studies, this study tries to hybridize SVM and rough set theory (RST), two commonly used credit risk analysis techniques, to create a novel hybrid intelligent mining system for credit risk evaluation. The main reason of selecting SVM and RST reflects the following two aspects. On the one hand, the two methods do not need any additional information about data like probability in statistics or membership in fuzzy set theory[28−29]. On the other hand, the two methods have the following complementary characteristics between them.

As is known to all, RST[30] is a powerful mathematical tool that handles vagueness and uncertainty and has been widely applied in many areas, such as data mining and knowledge discovery in databases. But the deterministic mechanism for the error description is rather simple in RST[31]. Therefore, the rules generated by RST are often unstable and have low classification accuracy[32].

Support vector machines (SVMs)[33] are a class of state-of-the-art classification algorithm that is known to be successful in a wide variety of applications because of their strong robustness, high generalization ability, and low classification error. However, an obvious drawback is that a SVM lacks explanation capability for their results when applied to classification problems, that is, the results obtained from SVM classifiers are not intuitive to humans and are hard to understand[34]. For example, when an unlabeled example is classified by the SVM classifier as positive or negative, the only explanation that can be provided is that the outputs of the variables are lower/higher than some threshold; such an explanation is completely non-intuitive to human experts[34]. Therefore, it is often difficult to extract rules from a trained SVM classifier.

In such situations, it is natural to hybridize RST and SVM due to their complementarities. One typical approach is to use RST as a preprocessing tool of SVM in order to accelerate or simplify the SVM training process for mining valuable knowledge from databases[35]. By eliminating the redundant data from databases, RST can greatly speed up the training time of SVM and improve its prediction accuracy. Although these hybrid approaches can get high prediction accuracy, some knowledge contained in SVM may still be incomprehensible for the users due to its low explanation capability.

Different from the previous studies, this study tries to design a novel hybrid intelligent mining system integrating RST and SVM from a new perspective. In the proposed hybrid intelligent mining system, the first step is to reduce the original data with RST from a two-dimensional reduction (2D-Reduction) view (attribute dimension and object dimension). The main goal of this step is to reduce the training burden and accelerate the learning process for the SVM processing later. Then, SVM is used to extract typical features and to filter its noises, thus further reducing the original data. Finally, the mined knowledge or classification rules are generated from the reduced data by RST, rather than from the trained SVM. Therefore, the main advantage of our proposed hybrid intelligent system is that it can overcome the difficulty of extracting rules from a training SVM and possess the robustness which is lacking for rough set based approaches. To illustrate the effectiveness of the proposed system, two publicly available credit datasets including both consumer credit and corporation credit are used.

The rest of the study is organized as follows. Some preliminaries about rough sets and support vector machine are described in Section 2. In Section 3, the proposed hybrid intelligent mining system incorporating SVM into RST is described and the algorithms to generate classification rules from reduced data are proposed. For illustration and verification purposes, two real-world credit datasets are used in Section 4. Finally, some conclusions are drawn in Section 5.

2 Preliminaries of Rough Set Theory and Support Vector Machine

2.1 Basic Concepts of Rough Set Theory

RST, first introduced by Pawlak[30,36], is a highly-competitive mathematical tool to deal with vagueness and uncertainty. The main idea of RST is that some information (data, knowledge) can be associated with every object. Objects characterized by the same information are indiscernible (similar) in term of the available information. The indiscernibility relation generated in this way is the mathematical basis of RST. The most important problems that can be solved by RST include: 1) finding description of sets of objects in terms of attribute values; 2) checking dependencies (full or partial) between attributes; 3) reducing attributes; 4) analyzing the significance of attributes; and 5) generating decision rules[37–38].

In RST, an approximation space P is an ordered pair $P = (U, R)$, where U is a set called a universe and R is a binary equivalence relation on U. The relation R is called an indiscernibility relation. Each member of U is termed an object and each equivalence class of R is called an elementary set. If two objects of universe belong to the same elementary set of R, then the two objects are indistinguishable. A finite union of elementary sets is called a definable set in P.

Let $A \subseteq U$ and E^* be a family of equivalence classes of R. Usually, the rough set A can be represented by its upper and lower approximations. The upper approximation of A in P is defined by$\overline{R}A = \{a \in U \,|[a]_R \cap A \neq \varPhi\}$, where $[a]_R$ is the equivalence class of R containing a. The lower approximation of A in P is defined by $\underline{R}A = \{a \in U \,|[a]_R \subseteq A\}$. The set $M_p(A) = \overline{R}A - \underline{R}A$ is called a boundary of A in P.

The representation of knowledge as an information system (i.e., decision table) is a main attraction of the rough sets and it is defined as follows.

An information system is a quadruple (U, Q, V, δ) in which U is a nonempty finite set of objects, o, Q is a finite set of attributes, $V = \bigcup\limits_{q \in Q} V_q$ and V_q is the domain of attribute q, δ is a mapping function such that $\delta(o, q) \in V_q$ for every $q \in Q$ and $o \in U$. Let S= (U, Q, V, δ) be an information system, $F \subseteq Q$, and $o_i, o_j \in U$. The objects o_i and o_j are indiscernible by set of attributes F in S, denoted by $o_i\widetilde{F}o_j$, iff for each f in F, $\delta(o_i, f)$= $\delta(o_j, f)$. $\widetilde{F}$ is an equivalence relation on U for every F in Q. The set of all equivalence classes of $\widetilde{F}$ is called a classification generated by F in S. An ordered pair $(U,\widetilde{F})$ is an approximate space P. For any $A \subseteq U$, the lower and upper approximations of A in P will be called F-lower and F-upper approximations of A in S and shown by $\underline{F}A$ and$\overline{F}A$.

The set of Q attributes may be viewed as a set of condition attributes (C) and a set of decision attributes (D). Let $\psi = (Y_1, Y_2, \cdots, Y_k)$ and $\psi' = (Y_1', Y_2', \cdots, Y_k')$ be the classifications generated by D and C in S, respectively. Y_i is an equivalence class of $\widetilde{D}$ and Y_i' is an equivalence class of $\widetilde{C}$. Let $L(C,\, D)$ be defined as $L(C,\, D) = \cup_{i=1}^{k} \underline{C} Y_i$. The degree of dependency between the condition attributes C and the decision attribute D is defined as $\gamma(C, D) = \frac{|L(C,D)|}{|U|}$. Let $F \subseteq C$ and $F' \subset F$. The set of attributes F is a reduct of attribute C if $\gamma(F, D) = \gamma(C, D)$ and $L(F', D) \neq L(C, D)$ for every F' in F.

2.2 Basic Ideas of Support Vector Machines

Support vector machine (SVM), a class of typical machine learning algorithm, is proposed by Vapnik[33]. SVM is a very specific type of learning algorithms characterized by the capacity control of the decision function, the use of kernel functions, and the sparsity of the solution[33]. Established on the unique theory of the structural risk minimization principle to estimate a function by minimizing an upper bound of the generalization error, SVM is resistant to the over-fitting problem and can model nonlinear relations in an efficient and stable way, thus becoming very popular in many fields. The basic idea of SVM is to maximize the margin between classes, which can be formulated as the following convex quadratic programming (QP) problem:

$$\begin{cases} \max \ \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j \\ \text{s.t.} \ \sum_{i=1}^{m} \alpha_i y_i = 0, \quad 0 \le \alpha_i \le C, \quad i = 1, 2, \cdots m, \end{cases} \tag{1}$$

where $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_m\}$ is a training set in R^d space, $\{y_1, y_2, \cdots, y_m\} \in \{-1,\ 1\}$ are the class label data, and α_i are Lagrange multipliers, C is a parameter that assigns penalty cost to misclassification of samples. By solving the above optimization problem, the form of decision function can be derived as

$$f(x) = \boldsymbol{w}^{\mathrm{T}} \boldsymbol{x} + b, \tag{2}$$

where $\boldsymbol{w} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i$ and b is a bias term. Only vectors corresponding to nonzero contribute to decision function, and called as support vectors. More details about SVM can refer to [33].

3 Proposed Hybrid Intelligent Mining System

3.1 General Framework of Hybrid Intelligent Mining System

Integrating the advantages of SVM and RST, this study proposes a novel hybrid intelligent mining system to extract some efficient classification rules for decision support. The general framework of the proposed hybrid intelligent mining system is illustrated in Figure 1.

As can be seen from Figure 1, the proposed hybrid intelligent mining system consists of four main procedures: object reduction by a rough set, attribute reduction by a rough set, feature selection by a support vector machine, and rule generation by a rough set. Because object reduction is a vertical reduction for original decision table and attribute reduction is a horizontal reduction for original decision table, the combination of the two reductions are named as two-dimensional reduction (2D-Reduction) and introduced by Hashemi et al.[39]. The detailed contents of every procedure are described in the subsequent subsections.

3.2 2D-Reductions by Rough Set

In an original information table, some irrelevant, redundant, and noisy features may affect the mining process and the final performance. Thus, the original decision table is required to

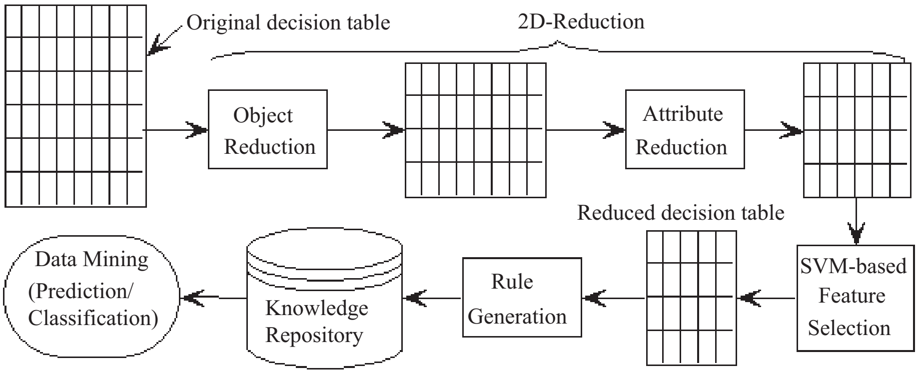

Figure 1 General framework of the hybrid intelligent mining system

be preprocessed. As previously mentioned, in rough sets an information table may not only be vertically reduced by removing some duplicate records (i.e., objects), but also be horizontally reduced by removing its redundant condition attributes. In a horizontally reduced information table, the condition attributes are independent and no attribute can be eliminated further without losing some information from the information table.

In the decision table, attributes deleted in the attribute reduction can be divided into two categories. One category contains irrelevant and redundant attributes that have no classification ability. An irrelevant attribute does not affect classification and a redundant feature does not add anything new to the classification[40]. The other category is noisy attributes. These attributes represent some classification ability, but this ability will disturb the mining of the true classification relation due to the effect of noise.

Many researchers have put forward their attribute reduction algorithms, and a comprehensive overview of these algorithms can be found in [40–41]. Recently, many methods based on rough sets[42−44] have been proposed for feature selection. However, in our study, the two-dimensional reduction (2D-Reduction) approach introduced by Hashemi et al.[39] is employed. In this approach, the information table is reduced both horizontally and vertically. The outcome of this reduction is called a 2D-Reduct. To define a 2D-reduced information table, both vertical and horizontal reductions are described as follows.

The vertical reduction of an information table S is $V=$ VERT(S) where V has all the properties of S, but it is free of duplicate records (i.e., objects).

The horizontal reduction of an information table S is $H=$ HORIZON(S) where: 1) H has all the objects of S; and 2) the set of condition attributes in H is the reduct of attributes in S[45].

The 2D-Reduction of an information table S is D^2 = 2D-RED(S) where D^2 = HORIZON (VERT(S)). A 2D-Reduct is generated using the 2D-Reduction algorithm, which is presented below.

2D-Reduction Algorithm.

Input: an information table, S. The set of condition attributes in S is C and the decision attribute is 2D. The condition attributes' value for object o_i in S is a set k_i and its decision is d_i.

Output: A minimal 2D-Reduct of S.

Step 1 Vertical reduction, i.e., removal of the reduction objects from S. For two objects, o_i and o_j in S: $(k_i = k_j)(d_i = d_j) \to (o_i = o_j)$. Thus, o_j is a redundant object and it will be removed from S.

Step 2 Horizontal reduction, i.e., removal of redundant condition attributes. Let V be the vertically reduced S, $V=$ VERT(S). Furthermore, let the number of objects in V be $n = |V|$, S and V have the same set of condition attributes. Let $P(C) = \{C^{(1)}, C^{(2)}, \cdots, C^{(m)}\}$ be the power set of C. Let S_i be an information system in which the conditions' set is $C^{(i)}$, ($C^{(i)} \neq C$ and $C^{(i)} \neq \phi$) and the objects are the same as objects in V. Let $V_i = \mathrm{VERT}(S_i)$, if $|V_i| = |V|$, then $C^{(i)}$ is a reduct of C. C may have several reducts $(C^{(1)}, C^{(2)}, \cdots, C^{(n)})$. If $C^{(i)} = \min(C^{(1)}, C^{(2)}, \cdots, C^{(n)})$, then $C^{(i)}$ is the minimal reduct of C and V_i is called a horizontal reduction of S. Also, V_i is called a minimal 2D reduct of S or simply a 2D-Reduct of S. Since C may have several minimal reducts, S may also have several 2D-Reducts.

In general, RST provides a useful technique to reduce irrelevant and redundant attributes from a large database with a lot of attributes[42−44]. But it is not so satisfactory for the reduction of noisy attributes because the classification region defined by RST is relatively simple. Furthermore, rough set based on attribute reduction criteria lacks an effective validation method[31]. Therefore, some new methods, such as neural networks and SVM, have been introduced to handle these noisy attributes[32]. Due to the fact that neural networks often suffer from over-fitting and local minima, SVM is used here.

3.3 Feature Selection by SVM

Feature selection, or attributes reduction, is a process of finding an optimal subset of all features according to some criteria so that the features subset is good enough to represent the classification relation of data. A good choice of the features subset provided to a classifier can increase its accuracy, save the computational time, and simplify its results[45].

As earlier noted, SVM[33] as a class of typical classification algorithm has the ability to deal with some complex classifications and possesses good robustness to noise, which gives us more chances to delete noisy attributes and reserve useful attributes. However, SVM is very inefficient for attribute reduction for their long learning time when dealing with a large number of attributes. This is a main reason why we do 2D-Reduction of original information using rough sets before using SVM.

The basic idea of SVM is to maximize the margin between classes, which can be formulated as a convex quadratic programming problem presented in Equation (1). From Equation (1), we can easily obtain its inverse-square of margin, which can be represented by

$$\|w\|^2 = \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j. \tag{3}$$

Feature selection methods that use the above quantity as a criterion have been proposed by several researchers[46−47]. For attribute reduction with SVM, it is possible to decompose the above quantity into sum of terms corresponding to original attributes or features:

$$\|w\|^2 = \sum_{k=1}^{d} \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j x_{ki} x_{kj}, \tag{4}$$

where x_{ki} is the kth component of $\boldsymbol{x}_i$. Therefore, contribution of kth feature to the inverse-square-margin can be given by

$$w_k^2 = \sum_{i,j=1}^{m} \alpha_i \alpha_j y_i y_j x_{ki} x_{kj}. \tag{5}$$

The importance of every feature was evaluated according to their values of w_k^2, and features having w_k^2 close to zero can be discarded without deteriorating classification performance.

Guyon et al.[46] employed an iterative procedure: starting with all features, a feature having the smallest w_k^2 was repeatedly set aside and w_k^2 was evaluated using parameter values of a classifier trained at each iteration. In this study, the iterative procedure is also used here due to simplicity.

3.4 Rule Generation by Rough Set

The reduced decision table obtained from the previous processes is free of noise and redundant information. By applying RST, the final knowledge, or classification rules are generated from the reduced decision table.

Actually, a reduced decision table can be seen as a rule set where each rule corresponds to one object of the table. The rule set can be generalized further by applying rough set value reduction method[32]. Its main idea is to drop those redundant condition values of rules and unite those rules in the same class.

Unlike most value reduction methods, which neglect the differences among the classification capabilities of condition attributes, our system first removes values of those attributes that have less discernibility degrees. In this way, more redundant values can be reduced from decision table and more concise rules can be generated. The steps of the rough-set-based rule generation algorithm are presented as follows.

Rough-set-based rule generation algorithm

Input: a reduced information table S. The set of conditions in S is C and the decision attribute is D. The conditions for object $o_i$in S are the set k_i and its decision is d_i.

Output: a rule set.

Step 1 Construct a discernibility relation matrix[48] and establish an initial rule set.

Step 2 Calculate the discernibility degree of condition attributes, sort the condition attributes by discernibility degree in an ascend way.

Step 3 Compare every rule in an initial rule set each other, and remove some redundant condition attribute values.

Step 4 Merge the same rules into one rule, transform the rules into more generalized rules for each class in D.

By applying RST, some classification rules are generated and can be used for data mining and knowledge discovery purposes.

3.5 General Procedure of the Hybrid Intelligent Mining System

Through the previous general framework and detailed process descriptions, a general procedure of the proposed hybrid intelligent mining system can be summarized as below.

Hybrid intelligent mining algorithm.

Step 1 Applying the 2D-Reduction algorithm, a 2D-Reduct can be obtained from an original decision table by vertical reduction and horizontal reduction.

Step 2 Comparing original decision table with 2D-Reduct, some attributes that are not in 2D-Reduct are removed from the original decision table and a reduced decision table can be obtained.

Step 3 Applying the feature selection approach of support vector machines, an important attributes subset can be obtained from 2D-Reduct.

Step 4 Comparing the reduced decision table with the important attribute subset, some attributes that are not in important attributes are removed. At the same time, remove some noisy objects and merge some identical objects into one object.

Step 5 Applying the rough set rule generation algorithm, the rule set or knowledge can be extracted from the reduced decision table.

4 Experiment Study

In this section, we explore the effectiveness of the proposed hybrid intelligent mining system with two real world credit datasets. For comparison, four individual classification models: logit regression (LogR), artificial neural network (ANN), rough set theory (RST), and support vector machine (SVM), and two hybrid classification models, rough neural network and fuzzy SVM, are used in the experiments. In addition, the classification accuracy in the testing set is used as performance evaluation criterion. Actually, three types of evaluation criteria are used to measure the classification results:

$$\text{Type I accuracy} = \frac{\text{number of both observed bad and classified as bad}}{\text{number of observed bad}}, \tag{6}$$

$$\text{Type II accuracy} = \frac{\text{number of both observed good and classified as good}}{\text{number of observed good}}, \tag{7}$$

$$\text{Total accuracy} = \frac{\text{number of correct classification}}{\text{the number of evaluation sample}}. \tag{8}$$

4.1 Corporation Credit Dataset

The corporation credit data about UK corporations come from the Financial Analysis Made Easy (FAME) CD-ROM database which can be found in the Appendix of [20]. It contains 30 failed and 30 non-failed firms. Twelve variables are used as the firms' characteristics description:

01) Sales;
02) ROCE: profit before tax/capital employed (%);
03) FFTL: funds flow (earnings before tax & depreciation)/total liabilities;
04) GEAR: (current liabilities + long-term debt)/ total assets;
05) CLTA: current liabilities/total assets;
06) CACL: current assets/current liabilities;
07) QACL: (current assets–stock)/current liabilities;
08) WCTA: (current assets − current liabilities)/ total assets;
09) LAG: number of days between account year end and the date the annual report and accounts were failed at company registry;
10) AGE: number of years the company has been operating since incorporation date;
11) CHAUD: coded 1 if changed auditor in previous three years, 0 otherwise;
12) BIG6: coded 1 if company auditor is a Big6 auditor, 0 otherwise.

In this experiment, 40 firms are randomly drawn as the training sample. Due to the scarcity of inputs, we make the number of good firms equal to the number of bad firms in both the training and testing samples, so as to avoid the embarrassing situations that just two or three good (or bad, equally likely) inputs in the testing samples. Thus, the training samples include 20 datasets in each class. This way of composing the sample of firms was also used by several researchers in the past, e.g., [1, 3, 19], among others. Its aim is to minimize the effect of such factors as industry or size that can be very important in some cases. Besides the above training samples, the testing samples were collected using a similar approach. The testing samples consist of 10 failed and 10 non-failed firms.

In the above 12 variables, the continuous variables or attributes are discretisized using the equal width bucket method for the rough set application. In the ANN model, a three-layer back-propagation neural network (BPNN) with 21 TANSIG neurons in the hidden layer and one PURELIN neuron in the output layer is used. The network training function is the TRAINLM. Besides, the learning rate is set to 0.15 and momentum rate 0.35. The accepted average squared error is 0.05 and the training epochs are 2000. In the SVM model, the regularization parameter is 10, and the kernel function is a radial basis function with $\sigma^2 = 1$. The above parameters are obtained from a trial and error method. For the two hybrid approaches, the design of rough neural network system is similar to that of [32] and the design of fuzzy SVM is identical to that used by Wang et al.[25]. Such a test is repeated 20 times and the final Type I, Type II, and Total accuracy are the average of the results of the 20 individual tests. According to the previous settings, the final computational results are shown in Table 1. Note that the values in brackets are standard deviations.

Table 1 Corporation credit risk evaluation results with different methods

Category	Model	Type I (%)	Type II (%)	Total (%)
Single	LogR	70.51 [5.47]	71.36 [6.44]	70.77 [5.96]
	ANN	72.14 [7.85]	74.07 [7.03]	73.63 [7.29]
	SVM	76.54 [6.22]	78.85 [5.51]	77.84 [5.82]
	RST	61.24 [4.54]	79.04 [5.45]	70.15 [4.97]
Hybrid	Fuzzy SVM	79.00 [5.98]	79.00 [5.32]	79.00 [5.65]
	Neuro-RST	77.32 [4.87]	83.54 [5.06]	80.38 [4.94]
	SVM-RST	81.45 [4.23]	85.43 [5.18]	83.37 [4.78]

In Table 1, we can observe the following conclusions.

1) Of the four single methods, the SVM performs much better than the RST and LogR. By a two-tail paired t-test, the difference between the SVM and RST/LogR is significant at 5% significance level, but the difference between SVM and ANN is insignificant at 5% significance level. Besides the reason that the SVM can overcome the local minima problem, other reasons are worth exploring further.

2) In the three hybrid approaches, the proposed hybrid SVM-RST performs the best, followed by the neural-RST and fuzzy SVM. The main reason is that the proposed hybrid intelligent mining system has integrated main advantages of RST and SVM, thus generating a synergetic effect. Related to Neuro-RST, the SVM can overcome the local minima problem, which is often occurred in neural networks.

3) The above two conclusions are made in view of the total accuracy. From the viewpoint of Type I and Type II accuracy, we find that the accuracy of Type II is higher than that of Type I. The main reason is that the classification of bad customers (Type I accuracy) is more difficult than that of good customers (Type II accuracy).

4) Although different measurements are used, the proposed hybrid intelligent mining system shows its advantages in terms of Type I, Type II, and Total accuracy. This implies that the proposed hybrid intelligent approach has potentials to corporation credit risk evaluation. The main reason is that hybridizing two diverse models can remedy the shortcomings of any individual method, thus increasing the classification accuracy.

4.2 Consumer Credit Dataset

The consumer credit data is about Australian credit card applications obtained from UCI Machine Learning Repository (http://www.ics.uci.edu/∼mlearn/databases/statlog/australian/). It consists of 307 good creditors and 383 bad creditors. Each sample is characterized by 14 attributes including 6 numerical and 8 categorical attributes. All attribute names and values have been changed to meaningless symbols to protect confidentiality of the data. 37 samples (5 percent) had one or more missing values. These missing values will be filled using the following rule. If the sample attribute is categorical type, the missing value will be filled by the mode of the attribute. If the sample attribute is continuous numerical type, the missing value will be filled by the mean of the attribute.

Similarly, some continuous variables or attributes are discretisized using the equal width bucket method. In the ANN model, a three-layer back-propagation neural network is used and has 14 input nodes, 16 TANSIG neurons in the hidden layer and one PURELIN neuron in the output layer. The network training function is the TRAINLM. In the SVM model, the regularization parameter is 50, and the kernel is a radial basis function with $\sigma^2 = 15$. The above parameters are obtained from trial and error. Such test is repeated 30 times and the final Type I, Type II, and Total accuracy are the average of the results of the 30 individual tests. Accordingly, the computational results are reported in Table 2. Note that the values in brackets are standard deviations.

Table 2 Consumer credit risk evaluation results with different methods

Category	Model	Type I (%)	Type II (%)	Total (%)
Single	LogR	58.75 [3.38]	67.76 [3.11]	62.78 [3.25]
	ANN	68.81 [3.42]	78.69 [3.93]	73.75 [3.66]
	SVM	74.59 [1.73]	88.44 [1.42]	81.51 [1.68]
	RST	71.96 [1.49]	89.03 [1.19]	80.44 [1.35]
Hybrid	Fuzzy SVM	81.32 [2.38]	92.81 [3.06]	87.04 [2.85]
	Neuro-RST	77.87 [3.87]	91.32 [2.68]	84.54 [3.09]
	SVM-RST	85.96 [1.61]	96.41 [1.23]	90.15 [1.47]

In Table 2, it is easy to draw some similar conclusions in comparison to Table 1. First, the performance of the SVM is the best of the four single methods. Second, the proposed hybrid intelligent mining approach consistently outperforms the comparable four individual methods and two hybrid approaches discussed in this study.

However, some conclusions that are different from Table 1 can also be found. Firstly, the total performance of SVM and RST is significantly better than that of LogR at 1% level. Furthermore, the difference between ANN and LogR is also significant at 5% level. This reveals that these intelligent approaches have an obvious advantage over the statistical method. Secondly, the performance of fuzzy SVM is better than that of Neuro-RST, which is different from the results in Table 1. In addition, the classification accuracy of the consumer credit is generally better than that of the corporation credit in terms of their performance and their standard deviations. The reasons leading to these phenomena are worth exploring in future studies.

To summarize, it is obvious that the performance of the proposed hybrid intelligent approach is better than those of the four individual approaches and two hybrid methods listed in the experiments. It correctly classifies 81.45% of bad instances and 85.43% of good firms and the total accuracy arrives at 83.37% in the corporation credit dataset. In the consumer credit dataset, it correctly classifies 96.41% of good instances and 85.96% of bad firms and the total accuracy reaches to 90.15%. The main reason for this advantage is two-fold. On the one hand, the proposed hybrid intelligent mining system achieves the synergy effect through the integration of SVM and RST. On the other hand, the proposed hybrid intelligent mining system can effectively filter the noises in the data, which makes the generated rules more accurate.

5 Conclusion Remarks

In this study, a hybrid intelligent mining system integrating rough set theory and support vector machine is designed to evaluate credit risks from a new perspective. Through rough set approach a decision table is first reduced by removing redundant attributes and duplicate objects with 2D-reduction algorithm. Then a SVM is trained to delete noisy attributes in the decision table. Finally, the classification rules are generated from the reduced decision table by rough sets.

To demonstrate the effectiveness of the proposed hybrid intelligent mining system, two publicly credit datasets are used. The experimental results obtained reveal that the proposed hybrid intelligent mining system can generate more accurate classification results than four individual methods (RST, SVM, ANN, and logit regression) and two hybrid approaches (Neuro-RST and fuzzy SVM). Such results indicate that the proposed hybrid intelligent mining system can be used as a very promising tool for credit risk evaluation and analysis. In addition, this proposed hybrid intelligent mining system can be extended to more applications in data mining and knowledge discovery fields.

References

[1] E. I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, 1968, **23**(3): 589–609.

[2] J. C. Wiginton, A note on the comparison of logit and discriminant models of consumer credit behaviour, *Journal of Financial Quantitative Analysis*, 1980, **15**(3): 757–770.

[3] C. V. Zavgren, Assessing the vulnerability to failure of American industrial firms: A logistic analysis, *Journal of Business Finance and Accounting*, 1985, **12**(1): 19–45.

[4] B. J. Grablowsky and W. K. Talley, Probit and discriminant functions for classifying credit applicants: A comparison, *Journal of Economic and Business*, 1981, **33**(3): 254–261.

[5] F. Glover, Improved linear programming models for discriminant analysis, *Decision Science*, 1990, **21**(4): 771–785.

[6] O. L. Mangasarian, Linear and nonlinear separation of patterns by linear programming, *Operations Research*, 1965, **13**(3): 444–452.

[7] W. E. Henley and D. J. Hand, A k-nearest-neighbour classifier for assessing consumer credit risk, *Statistician*, 1996, **45**(1): 77–95.

[8] P. Makowski, Credit scoring branches out, *Credit World*, 1985, **75**(1): 30–37.

[9] R. Smalz and M. Conrad, Combining evolution with credit apportionment: A new learning algorithm for neural nets, *Neural Networks*, 1994, **7**(2): 341–351.

[10] K. K. Lai, L. Yu, S. Y. Wang, and L. G. Zhou, Credit risk analysis using a reliability-based neural network ensemble model, *Lecture Notes in Computer Science*, 2006, **4132**: 682–690.

[11] R. Malhotra and D. K. Malhotra, Evaluating consumer loans using neural networks, *Omega*, 2003, **31**(1): 83–96.

[12] L. Yu, S. Y. Wang, and K. K. Lai, Credit risk assessment with a multistage neural network ensemble learning approach, *Expert Systems with Applications*, 2008, **34**(2): 1434–1444.

[13] M. C. Chen and S. H. Huang, Credit scoring and rejected instances reassigning through evolutionary computation techniques, *Expert Systems with Applications*, 2003, **24**(4): 433–441.

[14] F. Varetto, Genetic algorithms applications in the analysis of insolvency risk, *Journal of Banking and Finance*, 1998, **22**(10): 1421–1439.

[15] T. Van Gestel, B. Baesens, J. Garcia, and P. Van Dijcke, A support vector machine approach to credit scoring, *Bank en Financiewezen*, 2003, **2**(1): 73–82.

[16] Z. Huang, H. C. Chen, C. J. Hsu, W. H. Chen, and S. S. Wu, Credit rating analysis with support vector machines and neural networks: A market comparative study, *Decision Support Systems*, 2004, **37**(4): 543–558.

[17] K. K. Lai, L. Yu, S. Y. Wang, and L. G. Zhou, Credit risk evaluation with least square support vector machine, *Lecture Notes in Computer Science*, 2006, **4062**: 490–495.

[18] L. Yu, S. Y. Wang, K. K. Lai, and L. G. Zhou, *Bio-Inspired Credit Risk Analysis–Computational Intelligence with Support Vector Machines*, Springer-Verlag, Berlin, 2008.

[19] A. I. Dimitras, R. Slowinski, R. Susmaga, and C. Zopounidis, Business failure prediction using rough sets, *European Journal of Operational Research*, 1999, **114**(2): 263–280.

[20] M. J. Beynon and M. J. Peel, Variable precision rough set theory and data discretisation: An application to corporate failure prediction, *Omega*, 2001, **29**(6): 561–576.

[21] T. S. Lee, C. C. Chiu, C. J. Lu, and I. F. Chen, Credit scoring using the hybrid neural discriminant technique, *Expert Systems with Applications*, 2002, **23**(3): 245–254.

[22] S. Piramuthu, Financial credit-risk evaluation with neural and neurofuzzy systems, *European Journal of Operational Research*, 1999, **112**(2): 310–321.

[23] R. Malhotra and D. K. Malhotra, Differentiating between good credits and bad credits using neuro-fuzzy systems, *European Journal of Operational Research*, 2002, **136**(1): 190–211.

[24] B. S. Ahn, S. S. Cho, and C.Y. Kim, The integrated methodology of rough set theory and artificial neural network for business failure prediction, *Expert Systems with Applications*, 2000, **18**(1): 65–74.

[25] Y. Q. Wang, S. Y. Wang, and K. K. Lai, A new fuzzy support vector machine to evaluate credit risk, *IEEE Transactions on Fuzzy Systems*, 2005, **13**(6): 820–831.

[26] L. C. Thomas, A survey of credit and behavioral scoring: forecasting financial risk of lending to consumers, *International Journal of Forecasting*, 2002, **16**(2): 149–172.

[27] L. C. Thomas, R. W. Oliver, and D. J. Hand, A survey of the issues in consumer credit modelling research, *Journal of the Operational Research Society*, 2005, **56**(9): 1006–1015.

[28] M. Yahia, R. Mahmod, N. Sulaiman, and F. Ahmad, Rough neural expert systems, *Expert Systems with Applications*, 2000, **18**(2): 87–99.

[29] H. P. Nguyen, L. L. Phong, P. Santiprabhob, and B. De Baets, Approach to generation rules for expert systems using rough set theory, in *Proceedings of the Joint* 9*th IFSA World Congress and* 20*th NAFIPS International Conference* (ed. by M. H. Smith, W. A. Gruver, and William, H. L. O'Higgins), Vancouver, Canada, 2001, 877–882.

[30] Z. Pawlak, Rough sets, *International Journal of Computing and Information Science*, 1982, **11**(5): 341–356.

[31] J. Bazan, A. Skowron, and P. Synak, Dynamic reducts as a tool for extracting laws from decisions tables, in *Proceedings of the* 8*th International Symposium on Methodologies for Intelligent Systems* (ed. by G. Goos, J. Hartmanis, and J. van Leeuwen), Springer-Verlag, Berlin, 1994, 346–355.

[32] R. Li and Z. Wang, Mining classification rules using rough sets and neural networks, *European Journal of Operational Research*, 2004, **157**(2): 439–448.

[33] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[34] G. Fung, S. Sandilya, and R. B. Rao, Rule extraction from linear support vector machines via mathematical programming, in *Rule Extraction from Support Vector Machines* (ed. by D. Joachim), Springer, Berlin, 2008, 83–107.

[35] Y. Li and T. Fang, Study of forecasting algorithm for support vector machines based on rough sets, *Journal of Data Acquisition & Processing*, 2003, **18**(2): 199–203.

[36] Z. Pawlak, Rough classification, *International Journal of Man-Machine Studies*, 1984, **20**(5): 469–483.

[37] Z. Pawlak, Rough sets, in *Rough Sets and Data Mining* (ed. by T. Y. Lin and N. Cercone), Kluwer, Dordrecht, 1997, 3–8.

[38] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer, Dordrecht, 1991.

[39] R. R. Hashemi, F. R., Jelovsek, and M. Razzaghi, Developmental toxicity risk assessment: A rough set approach, *International Journal of Methods of Information in Medicine*, 1993, **32**(1): 47–54.

[40] M. Dash and H. Liu, Feature selection for classification, *Intelligent Data Analysis*, 1997, **1**(3): 131–156.

[41] R. W. Swiniarski and A. Skowron, Rough set methods in feature selection and recognition, *Pattern Recognition Letters*, 2003, **24**(6): 833–849.

[42] N. Zhong, J. Dong, and S. Ohsuga, Using rough sets with heuristics for feature selection, *Journal of Intelligent Information Systems*, 2001, **16**(2): 199–214.

[43] M. Beynon, Reducts within the variable precision rough sets model: A further investigation, *European Journal of Operational Research*, 2001, **134**(3): 592–605.

[44] R. Slowinski, C. Zopoundis, and A. I. Dimitras, Prediction of company acquisition in Greece by means of the rough set approach, *European Journal of Operational Research*, 1997, **100**(1): 1–15.

[45] M. Last, A. Kandel, and O. Maimon, Information-theoretic algorithm for feature selection, *Pattern Recognition Letters*, 2001, **22**(6–7): 799–811.

[46] I. Guyon, J. Weston, and S. Barnhill, Gene selection for cancer classification using support vector machines, *Machine Learning*, 2002, **46**(1–3): 389–422.

[47] J. Brank, M. Grobelnik, N. Milic-Frayling, and D. Mladenic, Feature selection using support vector machines, in *Proceeding of the 3th International Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields*, Bologna, Italy, 2002, 25–27.

[48] R. Felix and T. Ushio, Rule induction from inconsistent and incomplete data using rough sets, in *IEEE International Conference on Systems, Man, and Cybernetics*, Tokyo, 1999, **5**: 154–158.