

Research Article

A Corporate Credit Rating Model Using Support Vector Domain Combined with Fuzzy Clustering Algorithm

Xuesong Guo, Zhengwei Zhu, and Jia Shi

School of Public Policy and Administration, Xi'an Jiaotong University, Xi'an 710049, China

Correspondence should be addressed to Xuesong Guo, guoxues1@163.com

Received 11 February 2012; Revised 19 April 2012; Accepted 9 May 2012

Academic Editor: Wanquan Liu

Copyright © 2012 Xuesong Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Corporate credit-rating prediction using statistical and artificial intelligence techniques has received considerable attentions in the literature. Different from the thoughts of various techniques for adopting support vector machines as binary classifiers originally, a new method, based on support vector domain combined with fuzzy clustering algorithm for multiclassification, is proposed in the paper to accomplish corporate credit rating. By data preprocessing using fuzzy clustering algorithm, only the boundary data points are selected as training samples to accomplish support vector domain specification to reduce computational cost and also achieve better performance. To validate the proposed methodology, real-world cases are used for experiments, with results compared with conventional multiclassification support vector machine approaches and other artificial intelligence techniques. The results show that the proposed model improves the performance of corporate credit-rating with less computational consumption.

1. Introduction

Techniques of credit ratings have been applied by bond investors, debt issuers, and governmental officials as one of the most efficient measures of risk management. However, company credit ratings are too costly to obtain, because agencies including Standard and Poor's (S&P), and Moody's are required to invest lots of time and human resources to accomplish critical analysis based on various aspects ranging from strategic competitiveness to operational level in detail [1–3]. Moreover, from a technical perspective, credit rating constitutes a typical multiclassification problem, because the agencies generally have much more than two categories of ratings. For example, ratings from S&P range from AAA for the highest-quality bonds to D for the lowest-quality ones.

The final objective of credit rating prediction is to develop the models, by which knowledge of credit risk evaluation can be extracted from experiences of experts and to be applied in much broader scope. Besides prediction, the studies can also help users capture fundamental characteristics of different financial markets by analyzing the information applied by experts.

Although rating agencies take emphasis on experts' subjective judgment in obtaining ratings, many promising results on credit rating prediction based on different statistical and Artificial Intelligence (AI) methods have been proposed, with a grand assumption that financial variables extracted from general statements, such as financial ratios, contain lots of information about company's credit risk, embedded in their valuable experiences [4, 5].

Among the technologies based on AI applied in credit rating prediction, the Artificial Neural Networks (ANNs) have been applied in the domain of finance because of the ability to learn from training samples. Moreover, in terms of defects of ANN such as overfitting, Support Vector Machine (SVM) has been regarded as one of the popular alternative solutions to the problems, because of its much better performance than traditional approaches such as ANN [6–11]. That is, an SVM's solution can be globally optimal because the models seek to minimize the structural risk [12]. Conversely, the solutions found by ANN tend to fall into local optimum because of seeking to minimize the empirical risk.

However, SVM, which was originally developed for binary classification, is not naturally modified for multiclassification of many problems including credit ratings. Thus, researchers have tried to extend original SVM to multiclassification problems [13], with some techniques of multiclassification SVM (MSVM) proposed, which include approaches that construct and combine several binary classifiers as well as the ones that directly consider all the data in a single optimization formulation.

In terms of multiclassification in the domain of credit rating containing lots of data, current approaches applied in MSVM still have some drawbacks in integration of multiple binary classifiers as follows.

- (1) Some unclassifiable regions may exist if a data point belongs to more than one class or to none.
- (2) Training binary classifiers based on two-class SVM multiple times for the same data set often result in a highly intensive time complexity for large-scale problems including credit ratings prediction to improve computational consumption.

To overcome the drawbacks associated with current MSVM in credit rating prediction, a novel model based on support vector domain combined with kernel-based fuzzy clustering is proposed in the paper to accomplish multiclassification involved in credit ratings prediction.

2. Literature Review

2.1. Credit Rating Using Data Mining Techniques

Major researches applying data mining techniques for bond rating prediction can be found in the literature.

Early investigations of credit rating techniques mainly focused on the applicability of statistical techniques including multiple discriminant analysis (MDA) [14, 15] and logistic regression analysis (LRA) [16], and so forth, while typical techniques of AI including ANN

Table 1: Prior bond rating prediction using AI techniques.

Research	Number of categories	AI methods applied	Data source	Samples size
[20]	2	BP	U.S	30/17
[21]	2	BP	U.S	126
[22]	3	BP	U.S (S&P)	797
[17]	6	BP, RPS	U.S (S&P)	110/60
[23]	6	BP	U.S (S&P)	N/A
[24]	6	BP	U.S (Moody's)	299
[25]	5	BP with OPP	Korea	126
[26]	6	BP, RBF	U.S (S&P)	60/60
[27]	5	CBR, GA	Korea	3886
[28]	5	SVM	U.S (S&P)	N/A
[29]	5	BP, SVM	Taiwan, U.S	N/A

[17, 18] and case-based reasoning (CBR) [19], and so forth are applied in the second phase of research.

The important researches applying AI techniques in bond-rating prediction are listed in Table 1. In summary, the most prior ones accomplish prediction using ANN with comparison to other statistical methods, with general conclusions that neural networks outperformed conventional statistical methods in the domain of bond rating prediction.

On the other hand, to overcome the limitations such as overfitting of ANN, techniques based on MSVM are applied in credit rating in recent years. Among the models based on MSVM in credit rating, method of Grammar and Singer was early proposed by Huang et al., with experiments based on different parameters so as to find the optimal model [29]. Moreover, methodologies based on One-Against-All, One-Against-One, and DAGSVM are also proposed to accomplish S&P's bond ratings prediction, with kernel function of Gaussian RBF applied and the optimal parameters derived from a grid-search strategy [28]. Another automatic-classification model for credit rating prediction based on One-Against-One approach was also applied [30]. And Lee applied MSVM in corporate credit rating prediction [31], with experiments showing that model based on MSVM outperformed other AI techniques such as ANN, MDA, and CBR.

2.2. Multiclassification by Support Vector Domain Description

Support Vector Domain Description (SVDD), proposed by Tax and Duin in 1999 [32] and extended in 2004 [33], is a method for classification with the aim to accomplish accurate estimation of a set of data points originally. The methods based on SVDD differ from two or multiclass classification in that a single object type is interested rather than to be separated from other classes. The SVDD is a nonparametric method in the sense that it does not assume any particular form of distribution of the data points. The support of unknown distribution of data points is modeled by a boundary function. And the boundary is "soft" in the sense that atypical points are allowed outside it.

The boundary function of SVDD is modeled by a hypersphere rather than a hyperplane applied in standard SVM, which can be made with less constraints by mapping the data points to a high-dimensional space using methodology known as kernel trick, where the classification is performed.

SVDD has been applied in a wide range as a basis for new methodologies in statistical and machine learning, whose application in anomaly detection showed that the model based on it can improve accuracy and reduce computational complexity [34]. Moreover, ideas of improving the original SVDD through weighting each data point by an estimate of its corresponding density were also proposed [35] and applied in area of breast cancer, leukemia, and hepatitis, and so forth. Other applications including pump failure detection [36], face recognition [37], speaker recognition [38], and image retrieval [39] are argued by researchers.

The capability of SVDD in modeling makes it one of the alternative to large-margin classifiers such as SVM. And some novel methods applied in multiclass classification were proposed based on SVDD [40] combined with other algorithms such as fuzzy theories [41, 42] and Bayesian decision [36].

3. The Proposed Methodology

In terms of SVDD, which is a boundary-based method for data description, it needs more boundary samples to construct a closely fit boundary. Unfortunately, more boundary ones usually imply that more target objects have to be rejected with the overfitting problem arising and computational consumption increased. To accomplish multiclassification in corporate credit rating, a method using Fuzzy SVDD combined with fuzzy clustering algorithm is proposed in the paper. By mapping data points to a high-dimensional space by Kernel Trick, the hypersphere applied to every category is specified by training samples selected as boundary ones, which are more likely to be candidates of support vectors. After preprocessing using fuzzy clustering algorithm, rather than by original ones directly in standard SVDD [32, 33], one can improve accuracy and reduce computational consumption. Thus, testing samples are classified by the classification rules based on hyperspheres specified for every class. And the thoughts and framework of the proposed methodology can be illustrated in Figures 1 and 2, respectively.

3.1. Fuzzy SVDD

3.1.1. Introduction to Hypersphere Specification Algorithm

The hypersphere, by which SVDD models data points, is specified by its center \mathbf{a} and radius R . Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots)$ denote the data matrix with n data points and p variables, which implies that \mathbf{a} is p -dimensional while R is scalar. The geometry of one solution to SVDD in two dimensions is illustrated in Figure 3, where ω_i represents the perpendicular distance from the boundary to an exterior points \mathbf{x}_i . In terms of interior points, and the ones positioned on the boundary, ω_i is to be assigned as 0. Hence, ω_i can be calculated using the following equation:

$$\omega_i = \max\{0, \|\mathbf{x}_i - \mathbf{a}\| - R\}. \quad (3.1)$$

In the following, another closely related measure can be obtained in (3.2) in terms of exterior points

$$\xi_i = \|\mathbf{x}_i - \mathbf{a}\|^2 - R^2 \implies \|\mathbf{x}_i - \mathbf{a}\|^2 = R^2 + \xi_i. \quad (3.2)$$

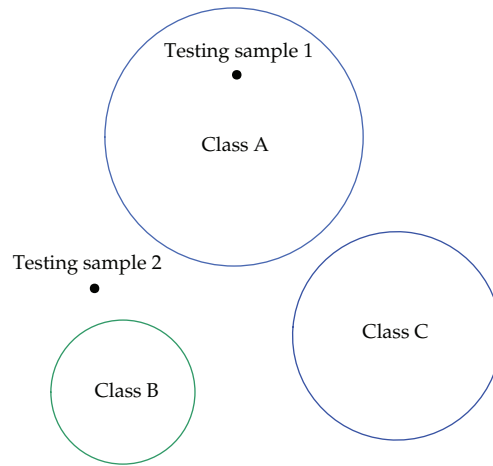


Figure 1: Multiclassification Based on SVDD.

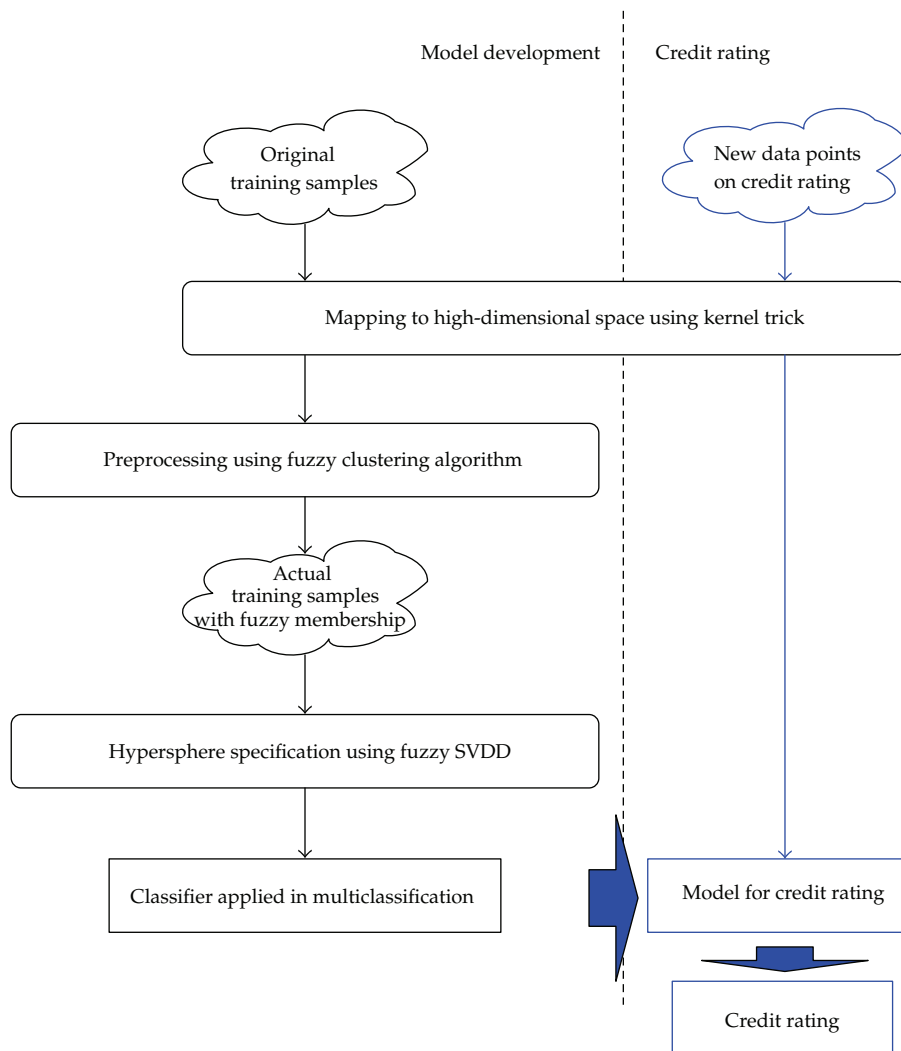


Figure 2: Framework of the Proposed Methodology.

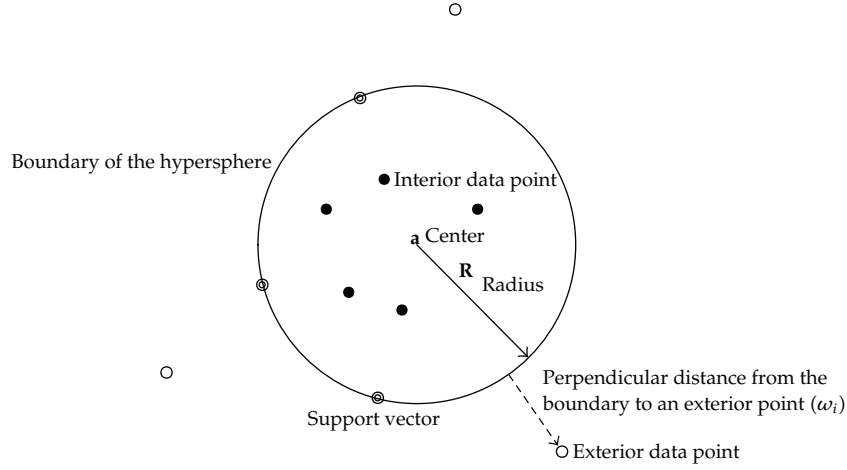


Figure 3: Geometry of the SVDD in two dimensions.

To obtain an exact and compact representation of the data points, the minimization of both the hypersphere radius and ξ_i to any exterior point is required. Moreover, inspired by fuzzy set theory, matrix \mathbf{X} can be extended to $\mathbf{X} = ((\mathbf{x}_1, s_1), (\mathbf{x}_2, s_2), (\mathbf{x}_3, s_3), \dots)$ with coefficients s_i representing fuzzy membership associated with \mathbf{x}_i introduced. So, the data domain description can be formulated as (3.3), where nonnegative slack variables ξ_i are a measure of error in SVDD, and the term $s_i \xi_i$ is the one with different weights based on fuzzy set theory

$$\begin{aligned} \min_{a, R, \xi} R^2 + C \sum_{i=1}^l s_i \xi_i, \\ \text{s.t. } \|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2 + \xi_i \quad \xi_i \geq 0, \quad i = 1, \dots, l. \end{aligned} \quad (3.3)$$

To solve the problem, the Lagrange Function is introduced, where $\alpha_i, \beta_i \geq 0$ are Lagrange Multipliers shown as follows:

$$L(R, a, \xi, \alpha, \beta) = R^2 + C \sum_{i=1}^l s_i \xi_i - \sum_{i=1}^l \alpha_i (R^2 + \xi_i - \|\mathbf{x}_i - \mathbf{a}\|^2) - \sum_{i=1}^l \beta_i \xi_i. \quad (3.4)$$

Setting (3.4) to 0, the partial derivatives of L leads to the following equations:

$$\begin{aligned} \frac{\partial L}{\partial R} &= 2R - 2R \sum_{i=1}^l \alpha_i = 0, \\ \frac{\partial L}{\partial \mathbf{a}} &= \sum_{i=1}^l \alpha_i (\mathbf{x}_i - \mathbf{a}) = 0, \\ \frac{\partial L}{\partial \xi_i} &= s_i C - \alpha_i - \beta_i = 0. \end{aligned} \quad (3.5)$$

That is,

$$\begin{aligned}\sum_{i=1}^l \alpha_i &= 1, \\ \mathbf{a} &= \sum_{i=1}^l \alpha_i \mathbf{x}_i, \\ \beta_i &= s_i C - \alpha_i.\end{aligned}\tag{3.6}$$

The Karush-Kuhn-Tucker complementarities conditions result in the following equations:

$$\begin{aligned}\alpha_i (R^2 + \xi_i - \|\mathbf{x}_i - \mathbf{a}\|^2) &= 0, \\ \beta_i \xi_i &= 0.\end{aligned}\tag{3.7}$$

Therefore, the dual form of the objective function can be obtained as follows:

$$L_D(\alpha, \beta) = \sum_{i=1}^l \alpha_i \mathbf{x}_i \cdot \mathbf{x}_i - \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j \mathbf{x}_i \cdot \mathbf{x}_j.\tag{3.8}$$

And the problem can be formulated as follows:

$$\begin{aligned}\max \quad & \sum_{i=1}^l \alpha_i \mathbf{x}_i \cdot \mathbf{x}_i - \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j \mathbf{x}_i \cdot \mathbf{x}_j \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq s_i C, \quad i = 1, 2, \dots, l, \\ & \sum_{i=1}^l \alpha_i = 1.\end{aligned}\tag{3.9}$$

The center of the hypersphere is a linear combination of data points with weighting factors α_i obtained by optimizing (3.9). And the coefficients α_i , which are nonzero, are thus selected as support vectors, only by which the hypersphere is specified and described. Hence, to judge whether a data point is within a hypersphere, the distance to the center should be calculated with (3.10) in order to judge whether it is smaller than the radius R . And the decision function shown as (3.12) can be concluded from

$$\left\| \mathbf{x} - \sum_{i=1}^l \alpha_i \mathbf{x}_i \right\|^2 \leq R^2,\tag{3.10}$$

$$R^2 = \left\| \mathbf{x}_{i0} - \sum_{i=1}^l \alpha_i \mathbf{x}_i \right\|^2 = \mathbf{x}_{i0} \cdot \mathbf{x}_{i0} - 2 \sum_{i=1}^l \alpha_i (\mathbf{x}_{i0} \cdot \mathbf{x}_i) + \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j \mathbf{x}_i \cdot \mathbf{x}_j,\tag{3.11}$$

$$\mathbf{x} \cdot \mathbf{x} - 2 \sum_{i=1}^l \alpha_i (\mathbf{x} \cdot \mathbf{x}_i) \leq \mathbf{x}_{i0} \cdot \mathbf{x}_{i0} - 2 \sum_{i=1}^l \alpha_i (\mathbf{x}_{i0} \cdot \mathbf{x}_i).\tag{3.12}$$

3.1.2. Introduction to Fuzzy SVDD Based on Kernel Trick

Similarly to the methodology based on kernel function proposed by Vapnik [12], the Fuzzy SVDD can also be generalized to high-dimensional space by replacing its inner products by kernel functions $K(\cdot, \cdot) = \Phi(\cdot) \bullet \Phi(\cdot)$.

For example, Kernel function of RBF can be introduced to SVDD algorithm, just as shown as follows:

$$\begin{aligned} \max & 1 - \sum_{i=1}^l \alpha_i^2 - \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j K(\mathbf{x}_i \cdot \mathbf{x}_j) \\ \text{s.t.} & 0 \leq \alpha_i \leq s_i C, \quad i = 1, 2, \dots, l, \\ & \sum_{i=1}^l \alpha_i = 1. \end{aligned} \quad (3.13)$$

And it can be determined whether a testing data point \mathbf{x} is within the hypersphere with (3.14) by introducing kernel function based on (3.12)

$$\sum_{i=1}^l \alpha_i K(\mathbf{x}, \mathbf{x}_i) \geq \sum_{i=1}^l \alpha_i K(\mathbf{x}_{i0}, \mathbf{x}_i). \quad (3.14)$$

3.2. Kernel-Based Fuzzy Clustering Algorithm

3.2.1. Introduction to Fuzzy Attribute C-Means Clustering

Based on fuzzy clustering algorithm [42], Fuzzy Attribute C-means Clustering (FAMC) [43] was proposed as extension of Attribute Means Clustering (AMC) and Fuzzy C-means (FCM).

Suppose $\chi \subset R^d$ denote any finite sample set, where $\chi = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$, and each sample is defined as $\mathbf{x}_n = (x_{1n}, x_{2n}, \dots, x_{dn})$ ($1 \leq n \leq N$). The category of attribute space is $F = \{C_1, C_2, \dots, C_c\}$, where c is the cluster number. For $\forall \mathbf{x} \in \chi$, let $\mu_{\mathbf{x}}(C_k)$ denote the attribute measure of \mathbf{x} , with $\sum_{k=1}^c \mu_{\mathbf{x}}(C_k) = 1$.

Let $\mathbf{p}_k = (\mathbf{p}_{k1}, \mathbf{p}_{k2}, \dots, \mathbf{p}_{kd})$ denote the k th prototype of cluster C_k , where $1 \leq k \leq c$.

Let μ_{kn} denote the attribute measure of the n th sample belonging to the k th cluster. That is, $\boldsymbol{\mu}_{kn} = \mu_n(\mathbf{p}_k)$, $\mathbf{U} = (\boldsymbol{\mu}_{kn})$, $\mathbf{P} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k)$. The task of fuzzy clustering is to calculate the attribute measure μ_{kn} , and determine the cluster which \mathbf{x}_n belongs to according to the maximum cluster index.

Fuzzy C-means (FCM) is an inner-product-induced distance based on the least-squared error criterion. A brief review of FCM can be found in Appendix based on coefficients definitions mentioned above.

Attribute Means Clustering (AMC) is an iterative algorithm by introducing the stable function [44]. Suppose $\rho(t)$ is a positive differential function in $[0, \infty)$. Let $\omega(t) = \rho'(t)/2t$, if $\omega(t)$, called as weight function, is a positive nonincreasing function, $\rho(t)$ is called as stable function. And $\rho(t)$ can be adopted as follows:

$$\rho(t) = \int_0^t 2s\omega(s)ds. \quad (3.15)$$

Hence, the relationship of objective function $\rho(t)$ and its weight function is described by stable function, which was introduced to propose AMC.

According to current researches, some alternative functions including squared stable function, Cauchy stable function, and Exponential stable function are recommended.

Based on previous researches, AMC and FCM are extended to FAMC, which is also an iterative algorithm to minimize the following objective function shown as (3.16), where $m > 1$, which is a coefficient of FCM introduced in Appendix

$$P(\mathbf{U}, \mathbf{p}) = \sum_{k=1}^c \sum_{n=1}^N \rho\left(\mu_{kn}^{m/2} \|\mathbf{x}_n - \mathbf{p}_k\|\right). \quad (3.16)$$

Moreover, procedure of minimizing (3.16) can be converted to an iterative objective function shown as (3.17) [43]

$$Q^{(i)}(\mathbf{U}, \mathbf{p}) = \sum_{k=1}^c \sum_{n=1}^N \omega\left(\left(\mu_{kn}^{(i)}\right)^{m/2} \|\mathbf{x}_n - \mathbf{p}_k^{(i)}\|\right) \left(\left(\mu_{kn}^{(i)}\right)^m \|\mathbf{x}_n - \mathbf{p}_k^{(i)}\|^2\right). \quad (3.17)$$

And the following equations can be obtained by minimizing $Q^{(i)}(\mathbf{U}^i, \mathbf{p})$, $Q^{(i)}(\mathbf{U}, \mathbf{p}^{(i+1)})$, respectively, which can be seen in [43, 45] in detail

$$\begin{aligned} \mathbf{p}_k^{(i+1)} &= \frac{\sum_{n=1}^N \omega\left(\left(\mu_{kn}^{(i)}\right)^{m/2} \|\mathbf{x}_n - \mathbf{p}_k^{(i)}\|\right) \left(\mu_{kn}^{(i)}\right)^m \mathbf{x}_n}{\sum_{n=1}^N \omega\left(\left(\mu_{kn}^{(i)}\right)^{m/2} \|\mathbf{x}_n - \mathbf{p}_k^{(i)}\|\right) \left(\mu_{kn}^{(i)}\right)^m}, \\ \mu_{kn}^{(i+1)} &= \frac{\omega\left(\left(\mu_{kn}^{(i)}\right)^{m/2} \|\mathbf{x}_n - \mathbf{p}_k^{(i)}\|\right) \left(\|\mathbf{x}_n - \mathbf{p}_k^{(i+1)}\|^2\right)^{-1/(m-1)}}{\sum_{k=1}^c \omega\left(\left(\mu_{kn}^{(i)}\right)^{m/2} \|\mathbf{x}_n - \mathbf{p}_k^{(i)}\|\right) \left(\|\mathbf{x}_n - \mathbf{p}_k^{(i+1)}\|^2\right)^{-1/(m-1)}}. \end{aligned} \quad (3.18)$$

3.2.2. Introduction to Kernel-Based Fuzzy Clustering

To gain a high-dimensional discriminant, FAMC can be extended to Kernel-based Fuzzy Attribute C-means Clustering (KFAMC). That is, the training samples can be first mapped into high-dimensional space by the mapping Φ using kernel function methods addressed in Section 3.1.2.

Since

$$\begin{aligned} \|\Phi(\mathbf{x}_n) - \Phi(\mathbf{p}_k)\| &= (\Phi(\mathbf{x}_n) - \Phi(\mathbf{p}_k))^T (\Phi(\mathbf{x}_n) - \Phi(\mathbf{p}_k)) \\ &= \Phi(\mathbf{x}_n)^T \Phi(\mathbf{x}_n) - \Phi(\mathbf{x}_n)^T \Phi(\mathbf{p}_k) - \Phi(\mathbf{p}_k)^T \Phi(\mathbf{x}_n) + \Phi(\mathbf{p}_k)^T \Phi(\mathbf{p}_k) \\ &= K(\mathbf{x}_n, \mathbf{x}_n) + K(\mathbf{p}_k, \mathbf{p}_k) - 2K(\mathbf{x}_n, \mathbf{p}_k) \end{aligned} \quad (3.19)$$

when Kernel function of RBF is introduced, (3.19) can be given as follows

$$\|\Phi(\mathbf{x}_n) - \Phi(\mathbf{p}_k)\|^2 = 2(1 - K(\mathbf{x}_n, \mathbf{p}_k)). \quad (3.20)$$

And parameters in KFAMC can be estimated by

$$\begin{aligned} \mu_{kn} &= \frac{(1 - K(\mathbf{x}_n, \mathbf{p}_k))^{-1/(m-1)}}{\sum_{k=1}^c (1 - K(\mathbf{x}_n, \mathbf{p}_k))^{-1/(m-1)}}, \\ p_k &= \frac{\sum_{n=1}^N \mu_{kn}^m K(\mathbf{x}_n, \mathbf{p}_k) \mathbf{x}_n}{\sum_{n=1}^N \mu_{kn}^m K(\mathbf{x}_n, \mathbf{p}_k)}, \end{aligned} \quad (3.21)$$

where $n = 1, 2, \dots, N, k = 1, 2, \dots, c$.

Moreover, the objective function of KFAMC can be obtained by substituting (3.16), (3.17) with (3.22), (3.23), respectively,

$$P(\mathbf{U}, \mathbf{p}) = \sum_{k=1}^c \sum_{n=1}^N \rho \left(\left\| \mu_{kn}^{m/2} (\Phi(\mathbf{x}_n) - \Phi(\mathbf{p}_k)) \right\| \right), \quad (3.22)$$

$$Q^{(i)}(\mathbf{U}, \mathbf{p}) = \sum_{k=1}^c \sum_{n=1}^N \omega \left(\left(\mu_{kn}^i \right)^{m/2} \left(1 - K(\mathbf{x}_n, \mathbf{p}_k^{(i)}) \right)^{1/2} \right) \left((\mu_{kn})^m (1 - K(\mathbf{x}_n, \mathbf{p}_k)) \right). \quad (3.23)$$

3.2.3. Algorithms of Kernel-Based Fuzzy Attribute C-Means Clustering

Based on theorem proved in [45], the updating procedure of KFAMC can be summarized in the following iterative scheme.

Step 1. Set c, m, ε and t_{\max} , and initialize $U^{(0)}, W^{(0)}$.

Step 2. For $i = 1$, calculate fuzzy cluster centers $P^{(i)}, U^{(i)}$. and $W^{(i)}$.

Step 3. If $|Q^{(i)}(U, P) - Q^{(i+1)}(U, P)| < \varepsilon$ or $i > t_{\max}$, stop, else go to Step 4.

Step 4. For step $i = i + 1$, update $P^{(i+1)}, U^{(i+1)}$, and $W^{(i)}$, turn to Step 3,

where i denotes iterate step, t_{\max} represents the maximum iteration times, and $W^{(i)}$ denotes the weighting matrix, respectively, which can be seen in [45] in detail.

3.3. The Proposed Algorithm

3.3.1. Classifier Establishment

In terms of SVDD, only support vectors are necessary to specify hyperspheres. But in the original algorithms [32, 33, 41], all the training samples are analyzed and thus computational cost is high consumption. Hence, if the data points, which are more likely to be candidates of support vectors, can be selected as training samples, the hypersphere will be specified with much less computational consumption.

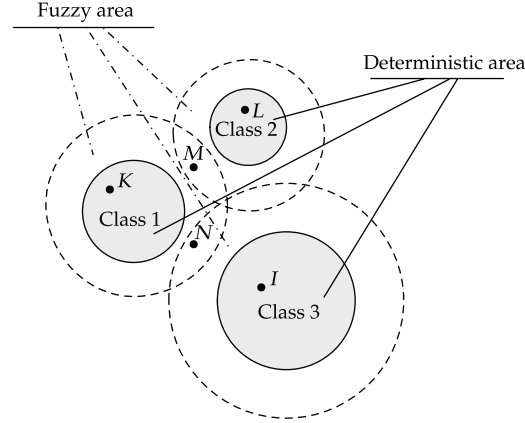


Figure 4: Thoughts of proposed methodology.

Just as illustrated in Figure 4, only the data points, such as M , N positioned in fuzzy areas, which are more likely to be candidates of support vectors, are necessary to be classified with SVDD, while the ones in deterministic areas can be regarded as data points belonging to certain class.

So, the new methodology applied in SVDD is proposed as follows.

- (1) Preprocess data points using FAMC to reduce amount of training samples. That is, if fuzzy membership of a data point to a class is great enough, the data point can be ranked to the class directly. Just as shown in Figure 5, the data points positioned in deterministic area (shadow area A) are to be regarded as samples belonging to the class, while the other ones are selected as training samples.
- (2) Accomplish SVDD specification with training samples positioned in fuzzy areas, which has been selected using KFAMC. That is, among the whole data points, only the ones in fuzzy area, rather than all the data points, are treated as candidates of support vectors. And the classifier applied in multiclassification can be developed based on Fuzzy SVDD by specifying hypersphere according to every class.

Hence, the main thoughts of Fuzzy SVDD establishment combined with KFAMC can be illustrated in Figure 6.

The process of methods proposed in the paper can be depicted as follows.

In high-dimensional space, the training samples are selected according to their fuzzy memberships to clustering centers. Based on preprocessing with KFAMC, a set of training samples is given, which is represented by $X_0^m = \{(x_1, \mu_1^m), (x_2, \mu_2^m), \dots, (x_l, \mu_l^m)\}$, where $l \in N$, $x_i \in R^n$, and $\mu_i^m \in [0, 1]$ denote the number of training data, input pattern, and membership to class m , respectively.

Hence, the process of Fuzzy SVDD specification can be summarized as follows.

Step 1. Set a threshold $\theta > 0$, and apply KFAMC to calculate the membership of each x_i , $i = 1, 2, \dots, l$, to each class. If $\mu_i^m \geq \theta$, μ_i^m is to be set as 1 and μ_i^t , $t \neq m$, is to be set as 0.

Step 2. Survey the membership of each x_i , $i = 1, 2, \dots, l$. If $\mu_i^m = 1$, x_i is to be ranked to class m directly and removed from the training set. And an updated training set can be obtained.

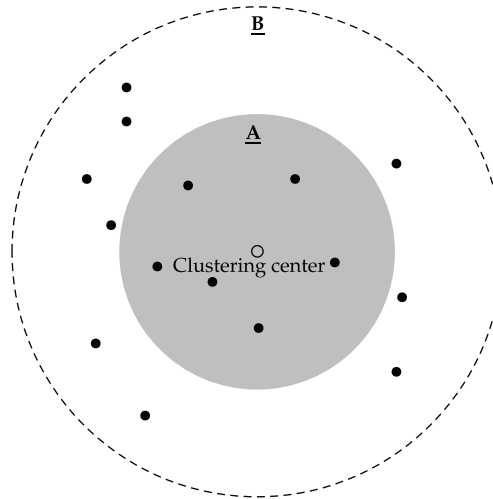


Figure 5: Data points selection using FAMC.

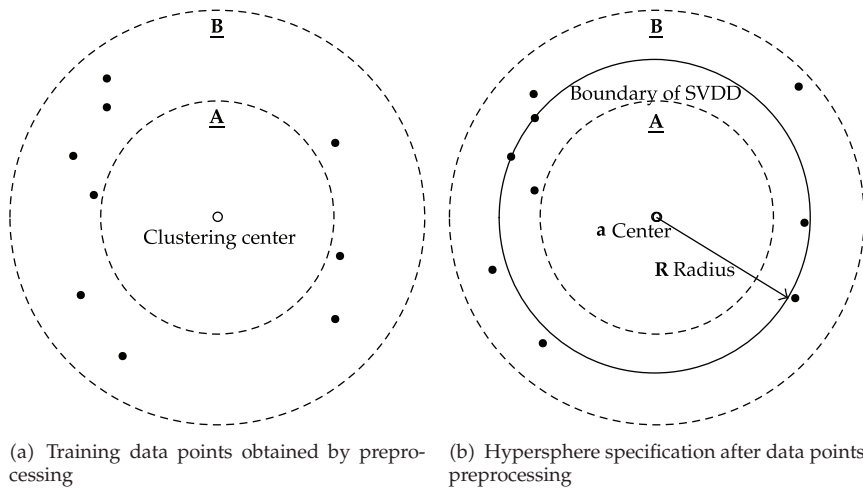


Figure 6: Fuzzy SVDD establishment.

Step 3. With hypersphere specified for each class using the updated training set obtained in Step 2, classifier for credit rating can be established using the algorithm of Fuzzy SVDD, just as illustrated in Figure 6.

3.3.2. Classification Rules for Testing Data Points

To accomplish multiclassification for testing data points using hyperspheres specified in Section 3.3.1, the following two factors should be taken into consideration, just as illustrated in Figure 7:

- (1) distances from the data point to centers of the hyperspheres;

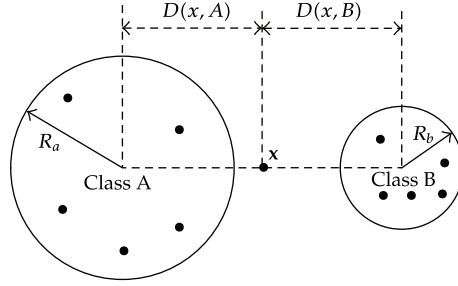


Figure 7: Classification of testing data point.

- (2) density of the data points belonging to the class implied with values of radius of each hypersphere.

Just as shown in Figure 7, $D(x, A)$, $D(x, B)$ denote the distances from data point x to center of class A and class B, respectively. Even if $D(x, A) = D(x, B)$, data point x is expected more likely to belong to class A rather than class B because of difference in distributions of data points. That is, data points circled by hypersphere of class A are sparser than the ones circled by hypersphere of class B since R_a is greater than R_b .

So, classification rules can be concluded as follows.

Let d denote the numbers of hyperspheres containing the data point.

Case I ($d = 1$). Data point belongs to the class represented by the hypersphere.

Case II ($d = 0$ or $d > 1$). Calculate the index of membership of the data point to each hypersphere using (3.24), where R_c denotes the radius of hypersphere c , $D(x_i, c)$ denotes the distance from data point x_i to the center of hypersphere c

$$\varphi(x_i, c) = \begin{cases} \lambda \left(\frac{1 - (D(x_i, c)/R_c)}{1 + (D(x_i, c)/R_c)} \right) + \gamma, & 0 \leq D(x_i, c) \leq R_c, \\ \gamma \left(\frac{R_c}{D(x_i, c)} \right), & D(x_i, c) > R_c, \end{cases} \quad \lambda, \gamma \in \mathbb{R}^+, \lambda + \gamma = 1. \quad (3.24)$$

And the testing data points can be classified according to the following rules represented with

$$F(x_i) = \arg \max_c \varphi(x_i, c). \quad (3.25)$$

4. Experiments

4.1. Data Sets

For the purpose of this study, two bond-rating data sets from Korea and China market, which have been used in [46, 47], are applied, in order to validate the proposed methodology. The data are divided into the following four classes: A1, A2, A3, and A4.

Table 2: Table of selected variables.

No.	Description	Definition
X1	Shareholders' equity	A firm's total assets minus its total liabilities
X2	Sales	Sales
X3	Total debt	Total debt
X4	Sales per employee	Sales/the number of employees
X5	Net income per share	Net income/the number of issued shares
X6*	Years after foundation	Years after foundation
X7	Gross earning to total asset	Gross earning/total Asset
X8	Borrowings-dependency ratio	Interest cost/sales
X9	Financing cost to total cost	Financing cost/total cost
X10	Fixed ratio	Fixed assets/(total assets-debts)
X11*	Inventory assets to current assets	Inventory assets/current assets
X12	Short-term borrowings to total borrowings	Short-term borrowings/total borrowings
X13	Cash flow to total assets	Cash flow/total assets
X14	Cash flow from operating activity	Cash flow from operating activity

*Indicates variables excluded in China data set.

4.2. Variables Selection

Methods including independent-samples t -test and F -value are applied in variable selection.

In terms of Korea data set, 14 variables, which are listed in Table 2, are selected from original ones, which were known to affect bond rating. For better comparison, similar methods were also used in China data set, with 12 variables among them being selected.

4.3. Experiment Results and Discussions

Based on the two data sets, some models based on AI are introduced for experiments. To evaluate the prediction performance, 10-fold cross validation, which has shown good performance in model selection [48], is followed. In the research, all features, which are represented with variables listed in Table 2, of data points range from 0 to 1 after Min-max transformation. To validate the methodology oriented multiclassification problem in credit rating, ten percent of the data points for each class are selected as testing samples. And the results of experiments on proposed method, with 0.9 being chosen as the value of threshold intuitively, are shown in Table 3.

To compare with other methods, the proposed model is compared with some other MSVM techniques, namely, ANN, One-Against-All, One-Against-One, DAGSVM, Grammer & Singer, OMSVM [46], and standard SVDD. The results concluded in the paper are all shown as average values obtained following 10-fold cross validation based on platform of Matlab 7.0.

To compare the performance of each algorithm, hit-ratio, which is defined according to the samples classified correctly, is applied. And the experiment results are listed in Table 4.

As shown in Table 4, the proposed method based on thoughts of hypersphere achieves better performance than conventional SVM models based on thoughts of hyperplane. Moreover, as one of modified models, some results obtained imply that the proposed method has better generalization ability and less computational complexity, which can be partially measured with training time labeled with "Time," than standard SVDD.

Table 3: Experimental results of the proposed method.

Data set No.	Korea data set		China data set	
	Train (%)	Valid (%)	Train (%)	Valid (%)
1	68.26	67.14	67.29	66.17
2	80.01*	71.23	68.35	67.13
3	73.21	70.62	71.56	71.01
4	75.89	72.37	75.24	72.36
5	76.17	74.23	84.17*	83.91*
6	75.28	75.01	80.02	79.86
7	78.29	76.23*	76.64	74.39
8	77.29	74.17	72.17	71.89
9	75.23	71.88	83.27	80.09
10	70.16	68.34	72.16	70.16
Avg.	74.98	72.12	75.09	73.70

*The best performance for each data set.

Table 4: Table of experiment results.

Type	Technique	Korea data set		China data set	
		Valid (%)	Time (Second)	Valid (%)	Time (Second)
Prior AI approach	ANN	62.78	1.67	67.19	1.52
	One-against-all	70.23	2.68	71.26	2.60
Conventional MSVM	One-against-one	71.76	2.70	72.13	2.37
	DAGSVM [28]	69.21	2.69	71.13	2.61
	Grammer & Singer [29]	70.07	2.62	70.91	2.50
	OMSVM [46]	71.61	2.67	72.08	2.59
The sphere-based classifier	Standard SVDD	72.09	1.70	72.98	1.04
	proposed method ($\theta = 0.9$)	72.12	1.20	73.70	0.86

Furthermore, as one of modified models based on standard SVDD, the proposed method accomplishes data preprocessing using KFAMC. Since the fuzzy area is determined by threshold θ , greater value of θ will lead to bigger fuzzy area. Especially, when $\theta = 1$, the algorithm proposed will be transformed to standard SVDD because almost all data points are positioned in fuzzy area. Hence, a model with too large threshold may be little different from standard SVDD, while a too small value will have poor ability of sphere-based classifier establishment due to lack of essential training samples. Thus, issues on choosing the appropriate threshold are discussed by empirical trials in the paper.

In the following experiment, the proposed method with various threshold values is tested based on different data sets, just as shown in Figure 8.

The results illustrated in Figure 8 showed that the proposed method achieved best performance with threshold of 0.9 based on Korea data set. But in terms of China market, it achieved best performance with the threshold of 0.8 rather than a larger one due to effects of more outliers existing in data set.

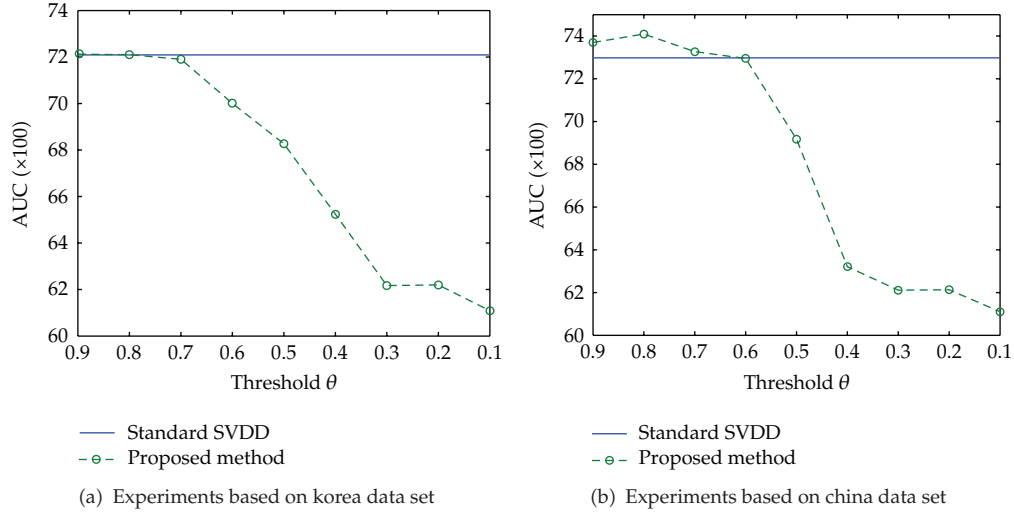


Figure 8: Experiment results of generalization ability on data sets. (AUC represents hit-ratio of testing samples).

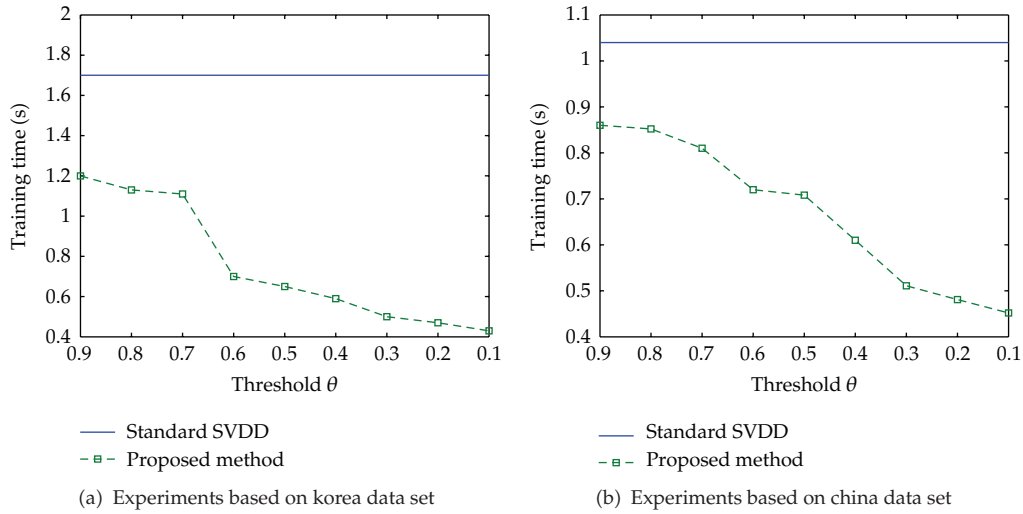


Figure 9: Experiment results of training time on data sets.

Moreover, training time of proposed method can be also compared with standard SVDD, just as illustrated in Figure 9.

Just as shown in Figure 9, results of experiments based on different data sets are similar. That is, with decline of threshold, more samples were eliminated from training set through preprocessing based on KFAMC to reduce training time. Hence, smaller values of threshold will lead to less computational consumption partly indicated as training time, while classification accuracy may be decreased due to lack of necessary training samples. Overall, threshold selection, which involves complex tradeoffs between computational consumption and classification accuracy, is essential to the proposed method.

5. Conclusions and Directions for Future Research

In the study, a novel algorithm based on Fuzzy SVDD combined with Fuzzy Clustering for credit rating is proposed. The underlying assumption of the proposed method is that sufficient boundary points could support a close boundary around the target data but too many ones might cause overfitting and poor generalization ability. In contrast to prior researches, which just applied conventional MSVM algorithms in credit ratings, the algorithm based on sphere-based classifier is introduced with samples preprocessed using fuzzy clustering algorithm.

As a result, through appropriate threshold setting, generalization performance measured by hit-ratio of the proposed method is better than that of standard SVDD, which outperformed many kinds of conventional MSVM algorithms argued in prior literatures. Moreover, as a modified sphere-based classifier, proposed method has much less computational consumption than standard SVDD.

One of the future directions is to accomplish survey studies comparing different bond-rating processes, with deeper market structure analysis also achieved. Moreover, as one of the MSVM algorithms, the proposed method can be applied in other areas besides credit ratings. And some more experiments on data sets such as *UCI* repository [49] are to be accomplished in the future.

Appendix

Brief Review of FCM

Bezdek-type FCM is an inner-product-induced distance-based least-squared error criterion nonlinear optimization algorithm with constrains,

$$J_m(U, P) = \sum_{k=1}^c \sum_{n=1}^N u_{kn}^m \|x_n - p_k\|_A^2,$$

$$\text{s.t. } U \in M_{fc} = \left\{ U \in R^{C \times N} \mid u_{kn} \in [0, 1], \forall n, k; \sum_{k=1}^c u_{kn} = 1, \forall n; 0 < \sum_{n=1}^N u_{kn} < N, \forall k \right\},$$
(A.1)

where u_{kn} is the measure of the n th sample belonging to the k th cluster and $m \geq 1$ is the weighting exponent. The distance between x_n and the prototype of k th cluster p_k is as follows:

$$\|x_n - p_k\|_A^2 = (x_n - p_k)^T A (x_n - p_k).$$
(A.2)

The above formula is also called as Mahalanobis distance, where A is a positive matrix. When A is a unit matrix, $\|x_n - p_k\|_A^2$ is Euclidean distance. We denote it as $\|x_n - p_k\|^2$ and

adopt Euclidean distance in the rest of the paper. So, the parameters of FCM are estimated by updating $\min J_m(\mathbf{U}, \mathbf{P})$ according to the formulas:

$$p_k = \frac{\sum_{n=1}^N (u_{kn})^m x_n}{\sum_{n=1}^N (u_{kn})^m},$$

$$u_{kn} = \frac{\|x_n - p_k\|^{-2/(m-1)}}{\sum_{i=1}^C \|x_n - p_i\|^{-2/(m-1)}}. \quad (\text{A.3})$$

Acknowledgment

The paper was sponsored by 985-3 project of Xi'an Jiaotong University.

References

- [1] A. Duff and S. Einig, "Understanding credit ratings quality: evidence from UK debt market participants," *The British Accounting Review*, vol. 41, no. 2, pp. 107–119, 2009.
- [2] W. F. Treacy and M. Carey, "Credit risk rating systems at large US banks," *Journal of Banking & Finance*, vol. 24, no. 1-2, pp. 167–201, 2000.
- [3] B. Becker and T. Milbourn, "How did increased competition affect credit ratings?" *Journal of Financial Economics*, vol. 101, no. 3, pp. 493–514, 2011.
- [4] B. Yang, L. X. Li, H. Ji, and J. Xu, "An early warning system for loan risk assessment using artificial neural networks," *Knowledge-Based Systems*, vol. 14, no. 5-6, pp. 303–306, 2001.
- [5] X. Zhu, H. Wang, L. Xu, and H. Li, "Predicting stock index increments by neural networks: the role of trading volume under different horizons," *Expert Systems with Applications*, vol. 34, no. 4, pp. 3043–3054, 2008.
- [6] H. Ahn, K. J. Kim, and I. Han, "Purchase prediction model using the support vector machine," *Journal of Intelligence and Information Systems*, vol. 11, pp. 69–81, 2005.
- [7] W. T. Wong and S. H. Hsu, "Application of SVM and ANN for image retrieval," *European Journal of Operational Research*, vol. 173, no. 3, pp. 938–950, 2006.
- [8] P. R. Kumar and V. Ravi, "Bankruptcy prediction in banks and firms via statistical and intelligent techniques—a review," *European Journal of Operational Research*, vol. 180, no. 1, pp. 1–28, 2007.
- [9] Y. Yang, "Adaptive credit scoring with kernel learning methods," *European Journal of Operational Research*, vol. 183, no. 3, pp. 1521–1536, 2007.
- [10] H. S. Kim and S. Y. Sohn, "Support vector machines for default prediction of SMEs based on technology credit," *European Journal of Operational Research*, vol. 201, no. 3, pp. 838–846, 2010.
- [11] G. Paleologo, A. Elisseeff, and G. Antonini, "Subagging for credit scoring models," *European Journal of Operational Research*, vol. 201, no. 2, pp. 490–499, 2010.
- [12] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer, New York, NY, USA, 1995.
- [13] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [14] G. E. Pinches and K. A. A. Mingo, "Multivariate analysis of industrial bond ratings," *The Journal of Finance*, vol. 28, no. 1, pp. 1–18, 1973.
- [15] A. Belkaoui, "Industrial bond ratings: a new look," *Financial Management*, vol. 9, no. 3, pp. 44–51, 1980.
- [16] L. H. Ederington, "Classification models and bond ratings," *The Financial Review*, vol. 20, no. 4, pp. 237–262, 1985.
- [17] J. W. Kim, H. R. Weistroffer, and R. T. Redmond, "Expert systems for bond rating: a comparative analysis of statistical, rule-based and neural network systems," *Expert Systems*, vol. 10, no. 3, pp. 167–172, 1993.
- [18] R. Chaveesuk, C. Srivaree-Ratana, and A. E. Smith, "Alternative neural network approaches to corporate bond rating," *Journal of Engineering Valuation and Cost Analysis*, vol. 2, no. 2, pp. 117–131, 1999.

- [19] H. J. Kim and K. S. Shin, "A hybrid approach using case-based reasoning and fuzzy logic for corporate bond rating," *Journal of Intelligence and Information system*, vol. 16, pp. 67–84, 2010.
- [20] S. Dutta and S. Shekhar, "Bond rating: a non-conservative application of neural networks," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 2, pp. 443–450, San Diego, Calif, USA, July 1988.
- [21] J. C. Singleton and A. J. Surkan, "Neural networks for bond rating improved by multiple hidden layers," in *Proceedings of the the IEEE International Conference on Neural Networks*, vol. 2, pp. 163–168, June 1990.
- [22] S. Garavaglia, "An application of a counter-propagation neural networks: simulating the standard & poor's corporate bond rating system," in *Proceedings of the 1st International Conference on Artificial Intelligence on Wall Street*, pp. 278–287, 1991.
- [23] J. Moody and J. Utans, "Architecture selection strategies for neural networks application to corporate bond rating," in *Neural Networks in the Capital Markets*, pp. 277–300, John Wiley & Sons, 1995.
- [24] J. J. Maher and T. K. Sen, "Predicting bond ratings using neural networks: a comparison with logistic regression," *Intelligent Systems in Accounting, Finance and Management*, vol. 6, no. 1, pp. 59–72, 1997.
- [25] Y. S. Kwon, I. G. Han, and K. C. Lee, "Ordinal pairwise partitioning (OPP) approach to neural networks training in bond rating," *Intelligent Systems in Accounting, Finance and Management*, vol. 6, no. 1, pp. 23–40, 1997.
- [26] R. Chaveesuk, C. S. Ratana, and A. E. Smith, "Alternative neural network approaches to corporate bond rating," *Journal of Engineering Valuation and Cost Analysis*, vol. 2, no. 2, pp. 117–131, 1999.
- [27] K. S. Shin and I. Han, "A case-based approach using inductive indexing for corporate bond rating," *Decision Support Systems*, vol. 32, no. 1, pp. 41–52, 2001.
- [28] L. Cao, L. K. Guan, and Z. Jingqing, "Bond rating using support vector machine," *Intelligent Data Analysis*, vol. 10, no. 3, pp. 285–296, 2006.
- [29] Z. Huang, H. Chen, C. J. Hsu, W. H. Chen, and S. Wu, "Credit rating analysis with support vector machines and neural networks: a market comparative study," *Decision Support Systems*, vol. 37, no. 4, pp. 543–558, 2004.
- [30] W. H. Chen and J. Y. Shih, "A study of Taiwan's issuer credit rating systems using support vector machines," *Expert Systems with Applications*, vol. 30, no. 3, pp. 427–435, 2006.
- [31] Y. C. Lee, "Application of support vector machines to corporate credit rating prediction," *Expert Systems with Applications*, vol. 33, no. 1, pp. 67–74, 2007.
- [32] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, no. 11–13, pp. 1191–1199, 1999.
- [33] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.
- [34] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 8, pp. 2282–2291, 2006.
- [35] D. Lee and J. Lee, "Domain described support vector classifier for multi-classification problems," *Pattern Recognition*, vol. 40, no. 1, pp. 41–51, 2007.
- [36] D. Tax, A. Ypma, and R. Duin, "Pump failure detection using support vector data descriptions," in *Proceedings of the 3rd International Symposium on Advances in Intelligent Data (IDA '99)*, pp. 415–425, 1999.
- [37] S. W. Lee, J. Park, and S. W. Lee, "Low resolution face recognition based on support vector data description," *Pattern Recognition*, vol. 39, no. 9, pp. 1809–1812, 2006.
- [38] X. Dong, W. Zhaohui, and Z. Wanfeng, "Support vector domain description for speaker recognition," in *Proceedings of the IEEE Signal Processing Society Workshop, Neural Networks for Signal Processing XI*, pp. 481–488, Falmouth, Mass, USA, 2001.
- [39] C. Lai, D. M. J. Tax, R. P. W. Duin, E. Pekalska, and P. Paclík, "A study on combining image representations for image classification and retrieval," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, no. 5, pp. 867–890, 2004.
- [40] K. Y. Lee, D. W. Kim, K. H. Lee, and D. Lee, "Density-induced support vector data description," *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 284–289, 2007.
- [41] L. L. Wei, W. J. Long, and W. X. Zhang, "Fuzzy data domain description using support vector machines," in *Proceedings of the 2nd International Conference on Machine Learning and Cybernetics*, pp. 3082–3085, November 2003.

- [42] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [43] J. Liu and M. Xu, "Bezdek type fuzzy attribute C-means clustering algorithm," *Journal of Beijing University of Aeronautics and Astronautics*, vol. 33, no. 9, pp. 1121–1126, 2007.
- [44] Q. S. Cheng, "Attribute means clustering," *Systems Engineering—Theory & Practice*, vol. 18, no. 9, pp. 124–126, 1998.
- [45] J. Liu and M. Xu, "Kernelized fuzzy attribute C-means clustering algorithm," *Fuzzy Sets and Systems*, vol. 159, no. 18, pp. 2428–2445, 2008.
- [46] K. J. Kim and H. Ahn, "A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach," *Computers & Operations Research*, vol. 39, no. 8, pp. 1800–1811, 2012.
- [47] S. Pang, *Credit Rating and Stock Market Prediction Model*, Science Press, Beijing, China, 2005.
- [48] S. M. Weiss and C. A. Kulikowski, *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning and Expert Systems*, Machine Learning, Morgan Kaufmann, San Mateo, Calif, USA, 1991.
- [49] C. L. Blake and C. J. Merz, "UCI repository of machine learning databases," Department of Information and Computer Sciences, University of California, Irvine, Calif, USA <http://www.ics.uci.edu/~mlearn/MLRepository.html>.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

