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.

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

[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 D AGSVM are also proposed to accomplish S&P’s bond ratings prediction, with kernel function of Gaussian RBF applied and the optimal parameters derived form 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 constrains 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 \( \mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots) \) 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 \( \omega_i \) 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, \( \omega_i \) is to be assigned as 0. Hence, \( \omega_i \) can be calculated using the following equation:

\[
\omega_i = \max\{0, \|\mathbf{x}_i - \mathbf{a}\| - R\}. \tag{3.1}
\]

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

\[
\xi_i = \|\mathbf{x}_i - \mathbf{a}\|^2 - R^2 \implies \|\mathbf{x}_i - \mathbf{a}\|^2 = R^2 + \xi_i. \tag{3.2}
\]
Figure 1: Multiclassification Based on SVDD.

Figure 2: Framework of the Proposed Methodology.
To obtain an exact and compact representation of the data points, the minimization of both the hypersphere radius and $\xi_i$ to any exterior point is required. Moreover, inspired by fuzzy set theory, matrix $X$ can be extended to $X = ((x_1, s_1), (x_2, s_2), (x_3, s_3), \ldots)$ with coefficients $s_i$ representing fuzzy membership associated with $x_i$ introduced. So, the data domain description can be formulated as (3.3), where nonnegative slack variables $\xi_i$ 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

$$
\min_{a, R, \xi} R^2 + C \sum_{i=1}^{l} s_i \xi_i,
$$

s.t. $||x_i - a||^2 \leq R^2 + \xi_i - \xi_i \geq 0, \quad i = 1, \ldots, l.$

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

$$
L(R, a, \xi, \alpha, \beta) = R^2 + C \sum_{i=1}^{l} s_i \xi_i - \sum_{i=1}^{l} \alpha_i \left( R^2 + \xi_i - ||x_i - a||^2 \right) - \sum_{i=1}^{l} \beta_i \xi_i.
$$

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

$$
\frac{\partial L}{\partial R} = 2R - 2R \sum_{i=1}^{l} \alpha_i = 0,
$$

$$
\frac{\partial L}{\partial a} = \sum_{i=1}^{l} \alpha_i (x_i - a) = 0,
$$

$$
\frac{\partial L}{\partial \xi_i} = s_i C - \alpha_i - \beta_i = 0.
$$
Mathematical Problems in Engineering

That is,

\[ \sum_{i=1}^{l} \alpha_i = 1, \]
\[ a = \sum_{i=1}^{l} \alpha_i x_i, \]
\[ \beta_i = s_i C - \alpha_i. \]

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

\[ \alpha_i \left( R^2 + \xi_i - \|x_i - a\|^2 \right) = 0, \]
\[ \beta_i \xi_i = 0. \]

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

\[ L_D(a, \beta) = \sum_{i=1}^{l} \alpha_i x_i \cdot x_i - \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j x_i \cdot x_j. \]  \hspace{1cm} (3.8)

And the problem can be formulated as follows:

\[ \max \sum_{i=1}^{l} \alpha_i x_i \cdot x_i - \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j x_i \cdot x_j \]
\[ \text{s.t.} \quad 0 \leq \alpha_i \leq s_i C, \quad i = 1, 2, \ldots, l, \]
\[ \sum_{i=1}^{l} \alpha_i = 1. \]  \hspace{1cm} (3.9)

The center of the hypersphere is a linear combination of data points with weighting factors \( \alpha_i \) obtained by optimizing (3.9). And the coefficients \( \alpha_i \), 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

\[ \left\| x - \sum_{i=1}^{l} \alpha_i x_i \right\|^2 \leq R^2, \]  \hspace{1cm} (3.10)
\[ R^2 = \left\| x_{i0} - \sum_{i=1}^{l} \alpha_i x_i \right\|^2 = x_{i0} \cdot x_{i0} - 2 \sum_{i=1}^{l} \alpha_i (x_{i0} \cdot x_i) + \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j x_i \cdot x_j, \]  \hspace{1cm} (3.11)
\[ x \cdot x - 2 \sum_{i=1}^{l} \alpha_i (x \cdot x_i) \leq x_{i0} \cdot x_{i0} - 2 \sum_{i=1}^{l} \alpha_i (x_{i0} \cdot x_i). \]  \hspace{1cm} (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 $K(\cdot, \cdot) = \Phi(\cdot) \cdot \Phi(\cdot)$.

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

$$
\max 1 - \sum_{i=1}^{l} \alpha_i^2 - \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j K(x_i \cdot x_j) \\
\text{s.t. } 0 \leq \alpha_i \leq s_i, \quad i = 1, 2, \ldots, l, \\
\sum_{i=1}^{l} \alpha_i = 1.
$$

(3.13)

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

$$
\sum_{i=1}^{l} \alpha_i K(x, x_i) \geq \sum_{i=1}^{l} \alpha_i K(x_0, x_i).
$$

(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 $\chi = \{x_1, x_2, \ldots, x_n\}$, and each sample is defined as $x_n = (x_{1n}, x_{2n}, \ldots, x_{dn})$ ($1 \leq n \leq N$). The category of attribute space is $F = \{C_1, C_2, \ldots, C_c\}$, where $c$ is the cluster number. For $\forall x \in \chi$, let $\mu_{s}(C_k)$ denote the attribute measure of $x$, with $\sum_{k=1}^{c} \mu_{s}(C_k) = 1$.

Let $p_k = (p_{k1}, p_{k2}, \ldots, p_{kd})$ denote the $k$th prototype of cluster $C_k$, where $1 \leq k \leq c$.

Let $\mu_{kn}$ denote the attribute measure of the $n$th sample belonging to the $k$th cluster. That is, $\mu_{kn} = \mu_{n}(p_k), U = (\mu_{kn}), p = (p_{1}, p_{2}, \ldots, p_{k})$. The task of fuzzy clustering is to calculate the attribute measure $\mu_{kn}$, and determine the cluster which $x_n$ 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:

$$
\rho(t) = \int_{0}^{t} 2s\omega(s) ds.
$$

(3.15)
Hence, the relationship of objective function $\rho(t)$ and its weight function is described by sable 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.

\[
P(U, p) = \sum_{k=1}^{c} \sum_{n=1}^{N} \rho \left( \mu_{kn}^{m/2} \| x_n - p_k \| \right).
\]  

(3.16)

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

\[
Q^{(i)}(U, p) = \sum_{k=1}^{c} \sum_{n=1}^{N} \omega \left( \left( \mu_{kn}^{(i)} \right)^{m/2} \| x_n - p_k^{(i)} \| \right) \left( \left( \mu_{kn}^{(i)} \right)^{m} \| x_n - p_k \|^2 \right).
\]  

(3.17)

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

\[
P_{k}^{(i+1)} = \frac{\sum_{n=1}^{N} \omega \left( \left( \mu_{kn}^{(i)} \right)^{m/2} \| x_n - p_k^{(i)} \| \right) \left( \left( \mu_{kn}^{(i)} \right)^{m} \| x_n - p_k \|^2 \right)}{\sum_{n=1}^{N} \omega \left( \left( \mu_{kn}^{(i)} \right)^{m/2} \| x_n - p_k^{(i)} \| \right) \left( \left( \mu_{kn}^{(i)} \right)^{m} \| x_n - p_k \|^2 \right)}.
\]  

(3.18)

\[
\mu_{kn}^{(i+1)} = \frac{\omega \left( \left( \mu_{kn}^{(i)} \right)^{m/2} \| x_n - p_k^{(i)} \| \right) \left( \| x_n - p_k^{(i+1)} \|^2 \right)^{-1/(m-1)}}{\sum_{k=1}^{c} \omega \left( \left( \mu_{kn}^{(i)} \right)^{m/2} \| x_n - p_k^{(i)} \| \right) \left( \| x_n - p_k^{(i+1)} \|^2 \right)^{-1/(m-1)}}.
\]

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 $\Phi$ using kernel function methods addressed in Section 3.1.2.

Since

\[
\| \phi(x_n) - \phi(p_k) \| = (\phi(x_n) - \phi(p_k))^T(\phi(x_n) - \phi(p_k))
\]

\[
= \phi(x_n)^T\phi(x_n) - \phi(x_n)^T\phi(p_k) - \phi(p_k)^T\phi(x_n) + \phi(p_k)^T\phi(p_k)
\]  

(3.19)

\[
= K(x_n, x_n) + K(p_k, p_k) - 2K(x_n, p_k)
\]
when Kernel function of RBF is introduced, (3.19) can be given as follows
\[
\|\Phi(x_n) - \Phi(p_k)\|^2 = 2(1 - K(x_n, p_k)).
\] (3.20)

And parameters in KFAMC can be estimated by
\[
\mu_{kn} = \frac{(1 - K(x_n, p_k))^{-1/(m-1)}}{\sum_{k=1}^{c} (1 - K(x_n, p_k))^{-1/(m-1)}},
\]
\[
p_k = \frac{\sum_{n=1}^{N} \mu_{kn}^m K(x_n, p_k) x_n}{\sum_{n=1}^{N} \mu_{kn}^m K(x_n, p_k)},
\] (3.21)

where \(n = 1, 2, \ldots, N, k = 1, 2, \ldots, c\).

Moreover, the objective function of KFAMC can be obtained by substituting (3.16), (3.17) with (3.22), (3.23), respectively,
\[
P(U, p) = \sum_{k=1}^{c} \sum_{n=1}^{N} \rho \left( \|\mu_{kn}^{m/2} (\Phi(x_n) - \Phi(p_k))\| \right),
\]
\[
Q^{(i)}(U, p) = \sum_{k=1}^{c} \sum_{n=1}^{N} \omega \left( (\mu_{kn}^{i})^{m/2} \left( 1 - K(x_n, p_k^{(i)}) \right)^{1/2} \right) ((\mu_{kn})^{m} (1 - K(x_n, p_k)))
\] (3.22) (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 \(c, m, \varepsilon\) and \(t_{\text{max}}\), 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 \(i > t_{\text{max}}\), 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, \(t_{\text{max}}\) 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 $\mathcal{X}_0^n = \{ (x_1, \mu_1^m), (x_2, \mu_2^m), \ldots, (x_l, \mu_l^m) \}$, where $l \in \mathbb{N}, x_i \in \mathbb{R}^n$, 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 $x_i, i = 1, 2, \ldots, l$, to each class. If $\mu_i^m \geq \theta$, $\mu_i^m$ is to be set as 1 and $\mu_t^i, t \neq m$, is to be set as 0.

**Step 2.** Survey the membership of each $x_i, i = 1, 2, \ldots, l$. 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;
(2) density of the data points belonging to the class implied with values of radius of each hypersphere.

Just as shown in Figure 7, \( D(x, A) \) and \( D(x, B) \) 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.

\textit{Case I (} \( d = 1 \)\textit{).} Data point belongs to the class represented by the hypersphere.

\textit{Case II (} \( d = 0 \text{ or } d > 1 \)\textit{).} 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 \( c \)

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

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

\[ F(x_i) = \arg \max_c \varphi(x_i, c). \] (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.

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

*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           |
| Conventional MSVM           | One-against-all         | 70.23          | 2.68           | 71.26          | 2.60           |
|                             | 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         | 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 $\theta$, greater value of $\theta$ 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,

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

s.t. \( U \in M_{fc} = \left\{ U \in \mathbb{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\}, \tag{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:

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

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 \( \min J_m(U, P) \) according to the formulas:

\[
\begin{align*}
    p_k &= \frac{\sum_{n=1}^{N} (u_{kn})^m x_n}{\sum_{n=1}^{N} (u_{kn})^m}, \\
    u_{kn} &= -\frac{\left\| x_n - p_k \right\|^{2/(m-1)}}{\sum_{i=1}^{C} \left\| x_n - p_k \right\|^{2/(m-1)}}.
\end{align*}
\] (A.3)

**Acknowledgment**

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

**References**


