Fault Diagnosis for Rolling Bearing Based on Semi-Supervised Clustering and Support Vector Data Description with Adaptive Parameter Optimization and Improved Decision Strategy

Jiawen Tan 1,2, Wenlong Fu 1,2,*, Kai Wang 1,2, Xiaoming Xue 3, Wenbing Hu 1,2 and Yahui Shan 4

1 College of Electrical Engineering & New Energy, China Three Gorges University, Yichang 443002, China; 201708520721060@ctgu.edu.cn (J.T.); 201808520721083@ctgu.edu.cn (K.W.); hwb@ctgu.edu.cn (W.H.)
2 Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station, China Three Gorges University, Yichang 443002, China
3 Faculty of Mechanical and Material Engineering, Huaiyin Institute of Technology, Huai’an 223003, China; taxueli@hyit.edu.cn
4 School of Hydropower and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China; shanyahui@hust.edu.cn
* Correspondence: fuwenlong@ctgu.edu.cn; Tel.: +86-136-1868-0447

Received: 28 March 2019; Accepted: 19 April 2019; Published: 23 April 2019

Abstract: Rolling bearing is of great importance in modern industrial products, the failure of which may result in accidents and economic losses. Therefore, fault diagnosis of rolling bearing is significant and necessary and can enhance the reliability and efficiency of mechanical systems. Therefore, a novel fault diagnosis method for rolling bearing based on semi-supervised clustering and support vector data description (SVDD) with adaptive parameter optimization and improved decision strategy is proposed in this study. First, variational mode decomposition (VMD) was applied to decompose the vibration signals into sets of intrinsic mode functions (IMFs), where the decomposing mode number K was determined by the central frequency observation method. Next, fuzzy entropy (FuzzyEn) values of all IMFs were calculated to construct the feature vectors of different types of faults. Later, training samples were clustered with semi-supervised fuzzy C-means clustering (SSFCM) for fully exploiting the information inside samples, whereupon a small number of labeled samples were able to provide sufficient data distribution information for subsequent SVDD algorithms and improve its recognition ability. Afterwards, SVDD with improved decision strategy (ID-SVDD) that combined with k-nearest neighbor was proposed to establish diagnostic model. Simultaneously, the optimal parameters C and σ for ID-SVDD were searched by the newly proposed sine cosine algorithm improved with adaptive updating strategy (ASCA). Finally, the proposed diagnosis method was applied for engineering application as well as contrastive analysis. The obtained results reveal that the proposed method exhibits the best performance in all evaluation metrics and has advantages over other comparison methods in both precision and stability.

Keywords: rolling bearing; fault diagnosis; variational mode decomposition; fuzzy entropy; semi-supervised clustering; k-nearest neighbor decision strategy; adaptive sine cosine algorithm; support vector data description

1. Introduction

Rolling bearing is one of the most commonly used components in mechanical equipment, whose running state directly affects the accuracy, reliability, and service life of the whole machine [1]. Hence,
how to recognize and diagnose rolling bearing faults remains one of the main concerns in preventing failures of mechanical systems [2]. However, rolling bearing is prone to failure due to the complex operating conditions, such as improper assembly, poor lubrication, water and foreign body invasion, corrosion or overload [3]. Therefore, effective methods need to be proposed to diagnosis a fault of rolling bearing, which can promote the reliability of industrial manufacture.

Owing to the rich information carried by vibration signals, most of fault diagnosis methods for rolling bearings rely on analyzing vibration signals [4,5], the resistance of the bearing can be measured using electrodynamic sensors or laser vibrometers [6]. Considering that vibration signals are commonly non-stationary, it is difficult to extract critical features directly from fault signals. To solve this problem, various non-stationary signal processing methods are proposed in previous studies. For instance, wavelet transform (WT) is a self-adaptive signal decomposition method [7], which inherits and develops the idea of localization of short-time Fourier transform and is an ideal tool for time-frequency analysis, but its generalization will be restricted once a wavelet basis is chosen. Empirical mode decomposition (EMD) has a strong ability to deal with non-stationary signals and does not need to pre-set any basis function [8], but its performance is affected by end effects and mode mixing. To overcome these defects, an improved version of EMD, ensemble empirical mode decomposition (EEMD), is proposed by introducing a noise-assisted analysis method [9], which is also increases the computational cost and cannot completely neutralize the added noise. Unlike the above methods, variational mode decomposition (VMD) is a novel quasi-orthogonal signal decomposition method which overcomes the generalization problem of WT or mode mixing and end effects of EMD and EEMD [10]. Its advancement and effectiveness have been demonstrated in previous studies [11,12]. With signals processed by non-stationary analytical method, the non-stationarity of signals would be weakened to some extent, and fault features can be further extracted. Due to the smoothness and robustness, as well as fewer samples required, fuzzy entropy (FuzzyEn) [13] is introduced to measure the features of vibration data and construct fault samples, which has been widely applied for feature extraction and shows superior performance in the field of fault diagnosis [14,15].

Fault diagnosis of rolling bearings is actually a problem of pattern recognition [16], and many machine learning methods have been developed for this problem [17]. Traditional machine learning methods require a large amount of labeled data to ensure the generalization of the algorithm. However, in practical problems, most of the samples are unlabeled, which results in unsatisfactory results of trained learning models in practical applications and makes a large number of unlabeled samples useless. Hence, semi-supervised methods are applied to fully exploit the information inside samples and expand the training set of machine learning algorithms. Semi-supervised fuzzy C-means clustering (SSFCM) [18] is a “semi-supervised version” of the fuzzy C-means clustering (FCM) algorithm. By introducing semi-supervised learning theory into FCM, training samples can be labeled under the supervision of the partly labeled samples, which makes up for the deficiency of labeled samples and improving the accuracy of pattern recognition [19,20]. In engineering application, several machine learning methods, including k-nearest neighbor (KNN) [21], artificial neural network (ANN) [22], support vector machine (SVM) [23] and support vector data description (SVDD) [24] are widely applied to solve pattern recognition problems. KNN is easy to realize and susceptible to the distribution of samples. AN has good performance in dealing with non-linear problems under the circumstance of a large number of samples. Based on structural risk minimization and statistical learning theory, SVM proposed by Vapnik finds an optimal hyperplane that meets the classification requirements, thus the hyperplane maximizes the interval between two classes while ensuring classification effect. Correspondingly, SVDD is a kernel method inspired by the idea of SVM theory [25], which possesses a prominent recognition ability through establishing a minimum hypersphere that contains as many samples as possible [26]. Currently, combined with relative distance decision strategy, SVDD has been successfully employed in the field of pattern recognition [27,28]. However, traditional relative distance decision strategy is difficult to accurately classify unknown samples locating in overlap regions of the hyperspheres and regions outside all the hyperspheres. To solve this problem, SVDD based on an
improved decision strategy (ID-SVDD) with the fusion of KNN method is proposed. To be specific, support vectors (SVs) are firstly extracted by establishing SVDD models. Next, the Euclidean distances between the testing sample and SVs are calculated. Whereupon, a fault type of the testing sample can be determined according to the proposed decision strategy.

Although SVDD has excellent effect in pattern recognition, its performance is affected by the parameters. In view of this, various optimization algorithms are proposed and applied to search the best parameters [29,30], including genetic algorithm (GA) [31], particle swarm optimization (PSO) [32], bacterial foraging algorithm (BFA) [33] and artificial sheep algorithm (ASA) [34]. As a novel optimization algorithm, sine cosine algorithm (SCA) proposed by Mirjalili [35] is proved to be effective in many studies [36,37]. To achieve better performance in convergence precision, a reformed sine cosine algorithm with adaptive strategy (ASCA) is developed in this paper. On the whole, a novel diagnosis method based on semi-supervised clustering and SVDD with adaptive parameter optimization and improved decision strategy is proposed in this study. Firstly, VMD is applied to decompose the vibration signals into sets of IMFs, and the decomposing mode number \( K \) is preset with central frequency observation method. Then FuzzyEn values of all IMFs are calculated to construct the feature vectors of corresponding samples with different fault types. Afterwards, training samples are clustered by SSFCM for information mining. Subsequently, SVDD with improved decision strategy (ID-SVDD) that combines with k-nearest neighbor is proposed to establish diagnostic model. Meanwhile, the optimal parameters \( C \) and \( \sigma \) for ID-SVDD is searched by the proposed adaptive sine cosine algorithm (ASCA). Finally, the engineering application and contrastive analysis indicate the availability and superiority of the proposed method.

The paper is organized as follows: Section 2 is dedicated to the basic knowledge of VMD, FuzzyEn, SSFCM, and SVDD. Section 3 introduces the proposed fault diagnosis method based on SSFCM and ASCA-optimized ID-SVDD. Section 4 presents the engineering application and comparative analysis, the experimental results of which demonstrate the superiority of the proposed method. Some discussion about the method presented in this paper is in Section 5. The conclusion is summarized in Section 6.

2. Fundamental Theories

2.1. Variational Mode Decomposition

Variational mode decomposition (VMD) is a novel non-recursive signal decomposition method [38], whose essence is to solve a variational optimization problem. Setting a scale \( K \) in advance, a given signal can be decomposed into \( K \) band-limited intrinsic mode functions (IMFs), among which each sub-signal is related to a certain center frequency. The constrained variation problem can be described as follows:

\[
\min_{m_k, \omega_k} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left( \delta(t) + \frac{i}{\pi t} \right) \cdot m_k(t) \cdot e^{-j\omega_k t} \right\|_2^2 \right\}
\]

s.t. \( \sum_{k=1}^{K} m_k = f, \quad k = 1, 2, \ldots, K \)

where \( m_k = \{m_1, m_2, \ldots, m_k\} \) represents the set of all mode functions, and \( \omega_k = \{\omega_1, \omega_2, \ldots, \omega_k\} \) represents the set of center frequencies, while \( \partial_t \) and \( \delta(t) \) are the partial derivative of time \( t \) for the function and unit pulse function, respectively. \( f \) is the given real valued input signal.

Lagrange multipliers are introduced to obtain the optimal solution of above constrained variational problem. Then problem (1) can be specified as follows:

\[
L(m_k, \omega_k, \beta) = \alpha \sum_{k=1}^{K} \left\| \partial_t \left( \delta(t) + \frac{i}{\pi t} \right) \cdot m_k(t) \cdot e^{-j\omega_k t} \right\|_2^2 + \left\| f(t) - \sum_{k} m_k(t) \right\|_2^2 + \left( \beta(t), f(t) - \sum_{k} m_k(t) \right)
\]

where \( \alpha \) and \( \beta(t) \) represent the penalty factor and Lagrange multiplier respectively.
Next, the alternate direction method of multipliers (ADMM) and iterative search are utilized to obtain the saddle point of Lagrange multipliers [39]. The optimization problems of $m_k$ and $\omega_k$ are formulated as (3) and (4) respectively.

$$m_k^{n+1} = \min \left\{ \alpha \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) m_k(t) e^{-j\omega t} \right] \right\|_2^2 + \left\| f(t) - \sum_i m_i(t) + \frac{\beta(t)}{2} \right\|_2^2 \right\}$$  \hspace{1cm} (3)

$$\omega_k^{n+1} = \min \left\{ \left\| \partial_t \left[ \left( \delta(t) + \frac{j}{\pi t} \right) m_k(t) e^{-j\omega t} \right] \right\|_2^2 \right\}$$  \hspace{1cm} (4)

Solving problems (3) and (4), the iterative equations are deduced as follows:

$$\hat{m}_k^{n+1}(\omega) = \frac{f(\omega) - \sum_{i \neq k} \hat{m}_i(\omega) + \frac{\hat{\beta}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}$$  \hspace{1cm} (5)

$$\omega_k^{n+1} = \frac{\int_{-\infty}^{\infty} \omega |\hat{m}_k(\omega)|^2 d\omega}{\int_{-\infty}^{\infty} |\hat{m}_k(\omega)|^2 d\omega}$$  \hspace{1cm} (6)

The Lagrange multipliers can be iterated with Equation (7).

$$\hat{\beta}^{n+1}(\omega) = \hat{\beta}^n(\omega) + \tau(f(\omega) - \sum_k \hat{m}_k^{n+1}(\omega))$$  \hspace{1cm} (7)

where $\tau$ is an updating parameter.

The main steps of VMD can be summarized as follows:

Step 1: Initialize $m_k^n, \omega_k^n, \beta^1, n = 1;

Step 2: Execute loop, $n = n + 1;

Step 3: Update $m_k$ and $\omega_k$ based on Equations (5) and (6);

Step 4: Update $\beta$ based on Equation (7);

Step 5: If $\sum_k \|\hat{m}_k^{n+1} - \hat{m}_k^n\|_2^2 / \|\hat{m}_k^n\|_2^2 < \epsilon$ loop end, else turn to step 2 for next iteration.

2.2. Fuzzy Entropy

To solve the defect that sample entropy [40] adopts a hard threshold criterion which may result in an unstable discrimination result, fuzzy entropy [41] introduces an exponential function from fuzzy theory as the fuzzy entropy function. For the time series $H = [h(1), h(2), ..., h(N)]$, set the fractal dimension $m$ and then construct $m$-dimensional vector. The obtained $m$-dimensional vector is as follows:

$$H_m(i) = [h(i), h(i+1), ..., h(i+m-1)] - g(i)$$  \hspace{1cm} (8)

where $i = 1, 2, ..., N-m+1$, and $g(i) = \frac{1}{m} \sum_{j=0}^{m-1} h(i+j)$.

Define the distance $d_{ij}$ between vector $H_m(i)$ and $H_m(j)$ as maximum absolute value of difference value between the corresponding element of them, that is:

$$d_{ij}^m = \max \left| h(i+k) - g(i) - [h(j+k) - g(j)] \right|$$  \hspace{1cm} (9)

where $i, j = 1, 2, ..., N-m+1$ and $i \neq j$; while $k = 0, 1, ..., m-1$. 
Introducing fuzzy membership function, the similarity between vector \( H_m(i) \) and \( H_m(j) \) can be defined as:

\[
A^m_{ij} = \begin{cases} 
1, & d^m_{ij} = 0 \\
\exp[-\ln2(\frac{d^m_{ij}}{r})^2], & d^m_{ij} > 0 
\end{cases}
\]  

(10)

where \( r \) is a positive real number, and \( r = R \cdot \text{std}(H) \), \( \text{std}(H) \) is the standard deviation of the given time series.

Define function \( \Phi^m(r) \) as:

\[
\Phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} C^m_i
\]  

(11)

where \( C^m_i = \frac{1}{N-m} \sum_{j \neq i} A^m_{ij} \).

Finally, FuzzyEn (with parameters \( m, N, r \)) of the time series can be deduced as the deviation of natural logarithm of \( \Phi^m \) from \( \Phi^{m+1} \), which means:

\[
FE(m, N, r) = \ln \Phi^m(r) - \ln \Phi^{m+1}(r)
\]  

(12)

2.3. Semi-Supervised Fuzzy C-Means Clustering

Semi-supervised fuzzy C-means clustering (SSFCM) [18] is an semi-supervised clustering algorithm based on the fuzzy C-means clustering (FCM) algorithm [42]. Specifically, supposing that the sample to be clustered is \( S = [s_1, s_2, \ldots, s_n] \) and the number of clusters is \( c \) (\( 2 \leq c \leq n \)), then the objective function of SSFCM is given as:

\[
J_{w,\alpha}(U, V; S, F) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^w d_{ij}^2 + \rho \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij} - f_{ij})^w d_{ij}^2
\]  

(13)

where \( U_{\text{cnn}} = [u_{ij}] \), \( u_{ij} \) represents the fuzzy membership of the \( i \)th sample belonging to the \( j \)th cluster, and the matrix formed by cluster centers is \( V = [v_1, v_2, \ldots, v_c]^T \). \( F_{\text{cnn}} = [f_{ij}], f_{ij} \) denotes the membership value of labeled samples, and \( b_{ij} \) is a two-valued (Boolean) indicator to distinguish labeled and unlabeled samples. Moreover, \( d_{ij} = ||s_i - v_j|| \) is the Euclidean distance between the \( i \)th sample and the \( j \)th cluster center, while \( w \) (\( w \geq 1 \)) is weighting parameter, and \( \rho (\rho \geq 0) \) is the balance coefficient.

With Lagrange multipliers introduced, the optimization problem of SSFCM can be converted into the following constrained minimization problem:

\[
J_{w,\alpha}(U, V, \lambda; S, F) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^w d_{ij}^2 + \rho \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij} - f_{ij})^w d_{ij}^2 - \sum_{i=1}^{c} \lambda_i \sum_{j=1}^{n} (u_{ij} - 1)
\]  

(14)

where \( \lambda_i, j = 1, 2, \ldots, n \), is the Lagrange multiplier. In the sequel, fuzzy membership \( u_{ij} \) and cluster center \( v_i \) can be iterated as:

\[
u_{ij} = \frac{1}{1 + \rho \left[ \frac{\sum_{k=1}^{c} d_{ij}^2 / d_{kj}^2}{\sum_{k=1}^{c} d_{ij}^2 / d_{kj}^2} + \rho f_{ij} \right]}
\]  

(15)

\[
v_i = \frac{\sum_{j=1}^{n} u_{ij}^2 b_{ij}}{\sum_{j=1}^{n} d_{ij}^2}
\]  

(16)

where \( i = 1, 2, \ldots, c, j = 1, 2, \ldots, n \).

On the whole, primary procedures of SSFCM can be summarized as follows:
Step 1: Initialize membership matrix \( U \) and give values of \( c, b, F \);
Step 2: Calculate prototypical cluster center \( V \);
Step 3: Update \( U \) based on Equation (15);
Step 4: Compare \( U' \) to \( U \), if \( \|U-U'\| < \delta \) (\( U' \) is the last iteration result and \( \delta \) is a certain tolerance) then loop end, else go to step 2.

2.4. Support Vector Data Description

Support vector data description (SVDD) [25] maps data set into a high-dimensional feature space and deduces a minimum hypersphere which contains as many target objects as possible. Given a target object set \( x_i, i = 1, 2, \ldots, n \), a hypersphere of SVDD can be established. Accordingly, the sphere of the hypersphere can be described by center \( a \) and radius \( R \) of the sphere. The expression for the hypersphere is as follows:

\[
\min f = R^2 + C \sum_i \xi_i
\]

\[s.t. \ |x_i - a|^2 \leq R^2 + \xi_i, \xi_i \geq 0, \forall i \quad (17)\]

where \( \xi_i \) is the slack variable, \( C \) is the penalty factor to reach a compromise between the size of hypersphere and number of misclassified samples. By introducing Lagrange multipliers, Equation (17) can be transformed into the following form:

\[
L = R^2 + C \sum_i \xi_i - \sum_i \alpha_i (R^2 + \xi_i - |x_i - a|^2) - \sum_i \beta_i \xi_i \quad (18)
\]

where \( \alpha \ (\alpha \geq 0) \) and \( \beta \ (\beta \geq 0) \) are Lagrange multipliers. Finally, a Gaussian radial basis function is applied to replace the inner product, transforming the optimization problem into the following dual problem:

\[
\max L = \sum_i \alpha_i K(x_i, x_j) - \sum_i \alpha_i \alpha_j K(x_i, x_j)
\]

\[K(x_i, x_j) = \exp\left(-\frac{|x_i - x_j|^2}{\sigma^2}\right) \quad (19)\]

\[s.t. \ \sum_i \alpha_i = 1, \ a = \sum_i \alpha_i x_i, 0 \leq \alpha_i \leq C, \ \forall i \]

where \( \sigma \) is the kernel parameter.

Solving programming problem (19), \( \alpha_i \) is obtained. To be specific, if \( 0 < \alpha_i < C \), the target sample is on the hypersphere, which is called the support vector (SV); if \( \alpha_i = 0 \), the sample is in the hypersphere; if \( \alpha_i = C \), the sample is outside the hypersphere, which is called bounded support vector (BSV). For a certain SV \( x_p \), radius of the hypersphere can be calculated as \( R \). For any sample \( y \) to be tested, its distance from the center of the hypersphere is \( D \):

\[
R = K(x_p, x_p) - 2 \sum_{i=1}^{n} \alpha_i K(x_i, x_p) + \sum_{i,j=1}^{n} \alpha_i \cdot \alpha_j K(x_i, x_j) \quad (20)
\]

\[
D = K(y, y) - 2 \sum_{i=1}^{n} \alpha_i K(x_i, y) + \sum_{i,j=1}^{n} \alpha_i \cdot \alpha_j K(x_i, x_j) \quad (21)
\]

3. Fault Diagnosis Based on Semi-Supervised Clustering and Support Vector Data Description with Adaptive Parameter Optimization and Improved Decision Strategy

3.1. Improved Decision Strategy

SVDD with relative distance decision strategy (RD-SVDD) is commonly applied for pattern recognition. However, it is difficult for relative distance decision strategy to achieve precise identification, especially for unknown samples within overlap regions of the hyperspheres and regions outside all the hyperspheres. The schematic diagram of positional relationship between
samples and hyperspheres is shown in Figure 1. Therefore, further development of decision strategy is required. To this end, SVDD with improved decision strategy (ID-SVDD) that fuses with the k-nearest neighbor (KNN) method is proposed in this research.

The KNN method proposed by Cover et al. [43] is one of the most widely used algorithms for pattern recognition, which is easy to understand and realize. The basic idea of KNN is to measure the differences between samples and their neighbors, and Euclidean and Manhattan distances are usually employed to achieve the measurement. Let $p, q$ be two samples with $m$ terms, the distances between them can be defined as:

\[
\text{Euclidean distance} : D(p, q) = \sqrt{\sum_{i=1}^{m} (p_i - q_i)^2}
\]

\[
\text{Manhattan distance} : D(p, q) = \sum_{i=1}^{m} |p_i - q_i|
\]

where $p_i$ and $q_i$ are the $i$th component of $p$ and $q$ respectively.

With training samples labeled by semi-supervised clustering, SVDD models of different types of faults are established, from which hypersphere radiuses $R$s and support vectors SVs can also be obtained. Furthermore, by comparing distance $D$ from testing sample to the centers of hyperspheres, it can be judged whether the sample is within the hypersphere. There are two cases after the comparison: If the sample to be classified is only in one hypersphere, the fault type of the sample is directly ascertained; otherwise, if the sample is in the overlap regions of the hyperspheres or regions outside all the hyperspheres, KNN is needed for further classification. To be specific, KNN firstly calculates the distances between testing sample and SVs, finding $k$-nearest neighbors among SVs whose distances are closest to the testing sample, then the type of the testing sample can be determined according to majority fault types of $k$-nearest neighbors.

The main steps of ID-SVDD are shown below:

Step 1: Establish SVDD models for different types of faults based on Equations (19) and (20), thus Rs and SVs are obtained;

Step 2: Calculate distance $D$ on the basis of Equation (21);

Step 3: Compare distance $D$ to the radius $R$ of each hypersphere, if testing sample is only in one hypersphere then go to step 6, else go to next step;

Step 4: Set KNN coefficient $k$, and find $k$-nearest neighbors according to Equation (22) or (23);

Step 5: Determine the type of testing sample with KNN decision;

Step 6: Decision completed.

The flowchart of ID-SVDD is shown in Figure 2.

![Figure 1: Schematic diagram of positional relationship between samples and hyperspheres.](image-url)
3.2. Adaptive Parameter Optimization

3.2.1. Sine Cosine Algorithm

Exploration and exploitation [35] are two phases in which the sine cosine algorithm (SCA) processes the optimization problem. In the phase of exploration, a set of random solutions are initiated as the outset of optimization process. With a strong randomness, SCA is able to search for feasible solutions quickly in the searching space. In the phase of exploitation, the random solutions change gradually, and the randomness is distinctly weaker than that of exploration phase, which is conducive to a better local searching.

Let \( Z_i = (Z_{i1}, Z_{i2}, \ldots, Z_{il})^T \) be the position of \( i \)th \((i = 1, 2, \ldots, M)\) individual, whereupon each solution of the optimization problem corresponds to the position of corresponding individual in the searching space, where \( l \) is the dimension of individuals. The best position of individual \( i \) is \( P_i = (P_{i1}, P_{i2}, \ldots, P_{il})^T \). During iterations, the position of individual \( i \) will be updated by the following Equations [35]:

\[
\begin{align*}
Z_{i1}^{k+1} &= Z_{i1}^k + r_1 \times \sin(r_2) \times |r_3 P_{i1}^k - Z_{i1}^k| \\
Z_{i2}^{k+1} &= Z_{i2}^k + r_1 \times \cos(r_2) \times |r_3 P_{i2}^k - Z_{i2}^k|
\end{align*}
\]

(24)

where \( Z_{ik}^k \) is the position of individual \( i \) in \( k \)th iteration, \( r_1, r_2, \) and \( r_3 \) are all random values.

The above equations are synthesized to:

\[
Z_{i1}^{k+1} = \begin{cases} 
Z_{i1}^k + r_1 \times \sin(r_2) \times |r_3 P_{i1}^k - Z_{i1}^k|, & r_4 < 0.5 \\
Z_{i1}^k + r_1 \times \cos(r_2) \times |r_3 P_{i1}^k - Z_{i1}^k|, & r_4 \geq 0.5
\end{cases}
\]

(25)

where \( r_4 \) is a random value within \( [0, 1] \).

As the above equations show, the updating equation has four main parameters, i.e., \( r_1, r_2, r_3 \) and \( r_4 \). The parameter \( r_1 \) contributes to determine the position’s region of individual \( i \) at next iteration. The parameter \( r_2 \) is a random number in the scope of \( [0, 2\pi] \), which defines distance that next movement should be towards. To stochastically emphasize (\( r_3 > 1 \)) or deemphasize (\( r_3 < 1 \)) the
effect of the best position of individual, a random weight $r_3$ within $[0, 2]$ is brought in the equations. Finally, the parameter $r_4$ is a random number in the range of $[0, 1]$ to switch fairly between sine and cosine components, specifically when $r_4 < 0.5$, sine component dominates iterations of the position of individual $i$, otherwise cosine component dominates iterations.

To seek feasible solutions in the searching space and eventually find the global optimum, exploration and exploitation phases should be in a balanced state. For this reason, the amplitudes of the sine and cosine functions are adaptively adjusted by changing $r_1$ as follows [35]:

$$r_1 = a - \frac{t}{T}$$  \hspace{1cm} (26)

where $T$, $t$ and $a$ are the maximum number of iterations, the current number of iterations and a constant respectively.

### 3.2.2. Adaptive Sine Cosine Algorithm

As mentioned in SCA, $r_1$ is a crucial parameter connecting with the sine and cosine functions $r_1\sin(r_2)$ or $r_1\cos(r_2)$. On the basis of the design principle of SCA, the algorithm firstly explores different regions and then exploits promising regions. Accordingly, the value of $r_1$ should be monotone decreasing during iterations. However, the updating strategy of $r_1$ as shown in Equation (26) is a linear decrease with equal-step, which will restrict the convergence and accuracy. For this purpose, an improved strategy for updating $r_1$ which is called adaptive sine cosine algorithm (ASCA) is proposed in this paper.

Considering that the fitness values of all individuals would change among the iterations, a hierarchical strategy based on the fitness values of individuals is introduced to divide all individuals into three subgroups [44]. More specifically, let $f_i$ be the fitness value of individual $i$, $f_a$ be the average fitness value of all individuals, and $f_e$ represents the average fitness value of individuals that fitness value better than $f_a$. If $f_i$ is better than $f_e$, individual $i$ is defined as elite individual; if $f_i$ is better than $f_a$ but worse than $f_e$, individual $i$ is defined as ordinary individual; if $f_i$ is worse than $f_a$, then define individual $i$ as inferior individual. The hierarchical strategy is illustrated in Figure 3a. After the above process, the number of individuals in different subgroups can be adaptively adjusted according to the fitness values during iterations, and different subgroups implement different adaptive operations. The updating strategy for $r_1$ is defined as follows and changes of $r_1$ for individuals of different subgroups is shown in Figure 3b:

1. **Elite individuals**: Are the best individuals among all ones, and close to the optimal value. A smaller $r_1$ is given with a cubic function to enhance local searching:

$$r_1 = (r_{\text{max}} - r_{\text{min}})(\frac{T - t}{T})^3 + r_{\text{min}}$$  \hspace{1cm} (27)

where $r_{\text{max}}$ and $r_{\text{min}}$ are the maximum and minimum value of $r_1$ respectively, while $T$ and $t$ are the maximum and current number of iterations respectively.

2. **Ordinary individuals**: Possess good global and local searching ability. $r_1$ maintains the original linear decreasing trend:

$$r_1 = a - \frac{t}{T}$$  \hspace{1cm} (28)

3. **Inferior individuals**: Are far from the optimal value. A bigger $r_1$ is needed to enhance global searching in early iteration, and to jump out the local optimum later:

$$r_1 = (r_{\text{min}} - r_{\text{max}})e^{-\frac{t}{b}} + r_{\text{max}}$$  \hspace{1cm} (29)

where $b$ is a positive constant to adjust exponential curve.
Figure 3. Schematic diagram of the proposed adaptive sine cosine algorithm (ASCA) based on hierarchical strategy: (a) Hierarchy of individuals (fitness value decreases from top down), (b) changes of $r_1$ for individuals of different subgroups.

3.3. Fault Diagnosis Based on SSFCM and ID-SVDD Optimized by ASCA

In this section, a novel model based on semi-supervised fuzzy C-means clustering, adaptive sine cosine algorithm and support vector data description with improved decision strategy (SSFCM-ASCA-ID-SVDD) is proposed to diagnose faults for rolling bearing. The specific steps are detailed as follows:

Step 1: Collect vibration signals;
Step 2: Determine the mode number $K$ by central frequency observation method;
Step 3: Decompose the signals into sets of IMFs with VMD;
Step 4: Calculate FuzzyEn values of all IMFs, and construct the fault feature vectors of training and testing samples;
Step 5: Cluster training samples by SSFCM, thus the training samples are all labeled;
Step 6: Optimize the parameters $C$ and $\sigma$ for SVDD with the proposed ASCA;
Step 7: Train the ID-SVDD model with the optimal parameters $C$ and $\sigma$;
Step 8: Apply the optimal ID-SVDD model to classify different types of faults and evaluate the performance of the model.

The flowchart of the proposed fault diagnosis model is shown in Figure 4.
Vibration signals $\x(t) = [x_1(t), x_2(t), \ldots, x_N(t)]$

Determine parameters $k$ of VMD

Decompose signal with VMD

IMF1...IMF$k$

Calculate FE values of each IMF

Labeled samples

Cluster FuzzyEn vectors by SSFCM

ID-SVDD model

Optimize ID-SVDD with ASCA

Classify fault samples and evaluate performance

Figure 4. The flowchart of the proposed fault diagnosis method.

4. Engineering Application

4.1. Data Collection

To validate the performance of the proposed method, a series of vibration signals with different fault locations and sizes were gathered from Bearings Data Center of Case Western Reserve University [45]. The experiment stand was mainly composed of a motor, an accelerometer, a torque sensor/encoder, and a dynamometer. The bearing was a deep groove ball bearing with model SKF6205-2RS. By using electro-discharge machining (EDM), the experiment device simulated three fault states of the rolling bearing: the inner race fault, ball element fault, and outer race fault. The depth of faults was 0.011 inches. Vibration signals collected from the drive end (DE) were taken as the research objects. In the experiment, the sample frequency was 12,000 Hz, and the rotation speed was 1750 rpm under the rated load of 2 hp. To fully verify the validity of the proposed fault diagnosis method, 9 types of samples were used in this paper, namely, the inner race fault, ball fault, and outer race fault with diameters of 0.007, 0.014, and 0.021 inches (i.e., each of the three types of faults has three defect sizes). Further, the vibration signal of each type of fault was divided into 59 segments containing 2048 sampling points. An image of the experiment device is shown in Figure 5. The experimental signals are listed in Table 1.
K with diameter of 0.007 inches (L1). As illustrated in Table 2, when taking different K as listed in Table 2, where K was ascertained by the sample of inner race fault with diameter of 0.007 inches (L1). As illustrated in Table 2, when K was set to 5, the first two center frequencies were relatively close, which meant that excessive decomposition occurred. The same conclusion can also be seen in Figure 6, during the iterative calculation of VMD, if K was 5 or greater, some center frequencies of IMFs would be relatively close to each other. For example, when K was 5, IMF1 and IMF2 were relatively close, while K was 6, IMF2 and IMF3 were relatively close. Therefore, parameter K was set to 4 in this research.

**Table 2.** Center frequencies with different K value.

<table>
<thead>
<tr>
<th>Number of Modes</th>
<th>Center Frequencies (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2903.44 1166.73</td>
</tr>
<tr>
<td>3</td>
<td>3832.19 2928.05 1285.99</td>
</tr>
<tr>
<td>4</td>
<td>3179.09 2236.65 1107.20 163.33</td>
</tr>
<tr>
<td>5</td>
<td>3945.87 3647.07 2922.65 1477.51 675.67</td>
</tr>
<tr>
<td>6</td>
<td>3952.42 3672.88 3042.01 2756.44 1465.45 668.06</td>
</tr>
<tr>
<td>7</td>
<td>4079.22 3835.39 3547.95 2921.58 1597.69 1263.47 633.97</td>
</tr>
<tr>
<td>8</td>
<td>4089.98 3848.10 3573.75 3040.86 2762.20 1583.65 1255.25 632.93</td>
</tr>
<tr>
<td>9</td>
<td>5175.26 4031.42 3821.29 3562.16 3040.24 2761.83 1583.05 1254.80 633.18</td>
</tr>
</tbody>
</table>
The decomposition results of samples with different fault types are illustrated in Figure 7, from which it can be seen that waveforms of different fault samples are different to some extent. After VMD decomposition, FuzzyEn values were calculated to construct the fault feature vectors, and the fractal dimension $m$ was set 2 while the positive real number $r$ was set as 0.2. The first three FuzzyEn vectors of different fault samples (L1–L9) are reported in Table 3.

![Figure 6. Center frequency distribution with different $K$ values.](image)

![Figure 7. The variational mode decomposition (VMD) decomposition results of signals with different faults: (a) inner race (0.007 inches); (b) inner race (0.014 inches); (c) inner race (0.021 inches); (d) ball (0.007 inches); (e) ball (0.014 inches); (f) ball (0.021 inches); (g) outer race (0.007 inches); (h) outer race (0.014 inches); (i) outer race (0.021 inches).](image)
Table 3. Fuzzy entropy values of different fault samples.

<table>
<thead>
<tr>
<th>Fault Label</th>
<th>Sample Number</th>
<th>IMF1</th>
<th>IMF2</th>
<th>IMF3</th>
<th>IMF4</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>1</td>
<td>1.7996</td>
<td>1.8052</td>
<td>1.1744</td>
<td>0.6738</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.7691</td>
<td>1.8104</td>
<td>1.1649</td>
<td>0.6638</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.8286</td>
<td>1.7767</td>
<td>1.1695</td>
<td>0.6738</td>
</tr>
<tr>
<td>L2</td>
<td>1</td>
<td>1.7242</td>
<td>1.3599</td>
<td>1.1695</td>
<td>0.5838</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.6353</td>
<td>1.2388</td>
<td>1.1776</td>
<td>0.5799</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.6377</td>
<td>1.3705</td>
<td>1.1981</td>
<td>0.6119</td>
</tr>
<tr>
<td>L3</td>
<td>1</td>
<td>2.0356</td>
<td>1.6703</td>
<td>1.8310</td>
<td>0.6125</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.9168</td>
<td>1.7186</td>
<td>1.8202</td>
<td>0.6199</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.9248</td>
<td>1.7081</td>
<td>1.8325</td>
<td>0.6094</td>
</tr>
<tr>
<td>L4</td>
<td>1</td>
<td>2.2160</td>
<td>2.0236</td>
<td>0.9334</td>
<td>0.3960</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.2282</td>
<td>2.0723</td>
<td>0.9072</td>
<td>0.3288</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.2025</td>
<td>2.0116</td>
<td>0.9650</td>
<td>0.5454</td>
</tr>
<tr>
<td>L5</td>
<td>1</td>
<td>1.7123</td>
<td>1.6451</td>
<td>1.2169</td>
<td>0.4758</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.9905</td>
<td>1.6487</td>
<td>1.1966</td>
<td>0.4806</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.9124</td>
<td>1.6921</td>
<td>1.1990</td>
<td>0.4936</td>
</tr>
<tr>
<td>L6</td>
<td>1</td>
<td>1.7338</td>
<td>2.0930</td>
<td>1.2618</td>
<td>0.6195</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.1333</td>
<td>1.9824</td>
<td>1.2164</td>
<td>0.6141</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.1229</td>
<td>1.9946</td>
<td>1.2435</td>
<td>0.6438</td>
</tr>
<tr>
<td>L7</td>
<td>1</td>
<td>1.2859</td>
<td>1.7536</td>
<td>1.5843</td>
<td>0.9124</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1.2313</td>
<td>1.7508</td>
<td>1.6109</td>
<td>0.8821</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1.3335</td>
<td>1.7556</td>
<td>1.6099</td>
<td>1.0993</td>
</tr>
<tr>
<td>L8</td>
<td>1</td>
<td>2.1004</td>
<td>2.0725</td>
<td>0.8502</td>
<td>0.4750</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2.1927</td>
<td>2.1651</td>
<td>0.7906</td>
<td>0.4937</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2.1565</td>
<td>2.1261</td>
<td>0.9207</td>
<td>0.5691</td>
</tr>
<tr>
<td>L9</td>
<td>1</td>
<td>0.7437</td>
<td>0.9596</td>
<td>0.9206</td>
<td>0.4920</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.6916</td>
<td>1.0609</td>
<td>1.2533</td>
<td>0.5486</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.9297</td>
<td>0.9185</td>
<td>0.9081</td>
<td>0.5329</td>
</tr>
</tbody>
</table>

The feature vectors of samples from different fault types were divided randomly into two parts with 40 and 19 vectors, respectively, the first part was used for training and the second one for testing. In order to better illustrate the characteristics and effects of SSFCM, training samples with ratio of 0.3 and 0.5 were randomly selected as labeled samples, while the rest were samples with unknown fault types. The weighting parameter \( w \) of SSFCM was set to 2 while the balance coefficient \( \alpha \) was set to \( 1/0.3 \) and \( 1/0.5 \), respectively, according to the ratio. Figure 8 depicts the 2D projection of training samples after clustering. The figure shows that training samples are gathered into nine clusters under the supervision of labeled samples and distributed near the center of each cluster. The contour-map displays the samples with the same partitioning values.

In the optimization phase, the proposed ASCA approach was applied to optimize the penalty factor \( C \) and the kernel parameter \( \sigma \), where the individual number was set to 30 and iteration times were set to 100. Considering the constraints of the optimization problem of SVDD, \( C \) should subject to \( C \geq 1/n \), where \( n \) is the number of target objects. Likewise, \( C \) should be subject to \( C < 1 \) or else it loses constraint on Lagrange multipliers. Taken together, searching ranges of \( C \) and \( \sigma \) were \([1/n, 1]\) and \([2^{-10}, 2^{10}]\), respectively. The coefficient \( k \) was set to 3 for KNN decision. During optimization process, the fitness value was calculated by five-fold cross validation on the training samples.

The optimization effect with the given \((C, \sigma)\) was measured by average accuracy of cross validation, then the optimal \((C, \sigma)\) was obtained. The convergence procedure of the adaptive parameter optimization is depicted in Figure 9a. It can be seen that the absolute average fitness value of all individuals increases rapidly at early stage of iterations and then tends to be stable, which means the
individuals were moving closer to the global optimal solution. The shaded part shows the distribution of the convergence curve in 10 optimization experiments. The comparison of ASCA and SCA is shown in Figure 9b, where all convergence experiments use the same FuzzyEn values and each convergence curve is the average of 10 experiments. The figure shows that both ASCA and SCA have a good convergence effect, but ASCA keeps a better convergence effect than SCA in the whole iterations. Besides, in the middle and later stages of iterations, the convergence curve of SCA decreases slightly, while ASCA remains stable.

![Figure 8. 2D projection of the training samples (0.5 labeled).](image)

![Figure 9. Convergence procedure of adaptive parameter optimization (0.5 labeled): (a) Distribution of convergence curves in 10 runs, (b) comparison of different optimization methods.](image)

With the optimal parameters \((C, \sigma)\) and labeled training samples, the ID-SVDD model was trained and applied to recognize testing samples. For an in-depth verification of the proposed method, diagnosis results were averaged over 10 runs and the training samples were chosen at random in every run. In this application, normalized mutual information (NMI) as well as accuracy (ACC) as two widely used evaluation metrics are employed to evaluate the diagnosis results [47], which reflects the matching degree between classification results and real sample distribution information. The scope of two metrics is both \([0, 1]\). The value closer to 1 indicates a better classification performance of the model. In addition, corresponding deviation scope for each result was also calculated as a reference for evaluation. The optimal \((C, \sigma)\) is presented corresponding to the best accuracy among 10 runs.
In contrastive experiments, parameters of CK-means, SCA, and RD-SVDD were set similar to the proposed method. That is, training samples with a ratio of 0.3 and 0.5 were randomly selected as labeled samples for CK-means clustering; SCA had 30 individuals and iterated 100 times; the ranges of $C$ and $\sigma$ of RD-SVDD were $[1/n, 1]$ and $[2^{-10}, 2^{10}]$ respectively. Considering the difference between SVDD and SVM, the searching scopes of parameter $C$ and $g$ of SVM were both set as $[2^{-10}, 2^{10}]$ respectively. Moreover, the optimal parameters $C$ and $\sigma (g)$ of all contrastive methods were selected in the same way the proposed method done. The performance of all contrastive methods was evaluated by NMI and ACC as well. A list of all important parameters in the experiment is shown in Table 4.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Method</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K$</td>
<td>mode number</td>
<td>VMD</td>
<td>4</td>
</tr>
<tr>
<td>$m$</td>
<td>fractal dimension</td>
<td>FuzzyEn</td>
<td>2</td>
</tr>
<tr>
<td>$r$</td>
<td>positive real number</td>
<td>FuzzyEn</td>
<td>0.2</td>
</tr>
<tr>
<td>$w$</td>
<td>weighting parameter</td>
<td>SFFCM</td>
<td>2</td>
</tr>
<tr>
<td>$a$</td>
<td>balance coefficient</td>
<td>SFFCM</td>
<td>1/0.3 and 1/0.5</td>
</tr>
<tr>
<td>$M$</td>
<td>number of individuals</td>
<td>ASCA</td>
<td>30</td>
</tr>
<tr>
<td>$T$</td>
<td>iteration times</td>
<td>ASCA</td>
<td>100</td>
</tr>
<tr>
<td>$C$</td>
<td>penalty factor</td>
<td>SVDD</td>
<td>0.1177 and 0.1523</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>kernel parameter</td>
<td>SVDD</td>
<td>26.7073 and 0.0011</td>
</tr>
<tr>
<td>$k$</td>
<td>number of nearest neighbors</td>
<td>KNN</td>
<td>3</td>
</tr>
</tbody>
</table>

The fault diagnosis results with different methods are shown in Table 5 and Figure 10. It can be seen from Table 5 that the proposed SSFCM-ASCA-ID-SVDD method achieves the best performance in terms of NMI and ACC metrics among all methods, i.e., 0.0868 and 0.9333 when the ratio is 0.3, 0.9254 and 0.0964 when the ratio is 0.5. To be specific, in the case of the ratio of 0.5: The comparison of FCM-ASCA-ID-SVDD, CK-means-ASCA-ID-SVDD and the proposed method indicates that NMI value of proposed method is 0.2913 and 0.0377 higher than other two methods, as well as that ACC value is 0.1228 and 0.0263 higher than other two methods, which shows SSFCM possess a better clustering precision. Similarly, the comparison analysis between SSFCM-SCA-ID-SVDD and the proposed method shows that ASCA-optimized ID-SVDD improves NMI value by 0.0152 and ACC value by 0.0099 than SCA-optimized ID-SVDD, indicating the effectiveness of the proposed adaptive parameter optimization strategy. Furthermore, the contrast among SSFCM-ASCA-SVM, SSFCM-ASCA-RD-SVDD and the proposed method shows that the NMI evaluation of proposed method is 0.0335 and 0.0568 higher than that of other two methods, while the ACC evaluation of proposed method is 0.0263 and 0.0409 higher than that of other two methods. Thus, it can be concluded that the proposed ID-SVDD model with improved decision strategy has better diagnosis performance.

Additionally, through the comprehensive analysis with a labeled sample ratio of 0.3 and 0.5, it showed that the proposed method achieved the best results under both the two ratio conditions. Meanwhile, the metrics under a ratio of 0.5 was better than that under a ratio of 0.3, which manifests that the clustering performance of SSFCM would be improved with the increasing of the labeled ratio, thus improving the diagnostic accuracy. Figure 11 is the boxplots of evaluation values, illustrating the performance of different diagnosis methods. As shown in the figure, the proposed SSFCM-ASCA-ID-SVDD method achieves a higher precision and possess better stability than other contrastive methods.
SSFCM-ASCASVM (0.3) 39.8545 3.4262 0.8865, [−0.030, 0.042] 0.9331, [−0.029, 0.024]
CK-means-ASCASVDD (0.3) 0.0983 72.2803 0.8834, [−0.029, 0.027] 0.9035, [−0.020, 0.032]
SSFCM-ASCASRD-SVDD (0.3) 0.0323 1024.0000 0.8505, [−0.069, 0.060] 0.9053, [−0.063, 0.054]
SSFCM-SCASIDA-SVDD (0.3) 0.9560 29.5815 0.8833, [−0.065, 0.033] 0.9333, [−0.056, 0.026]
SSFCM-ASCASID-SVDD (0.3) 0.1177 26.7073 0.8868, [−0.050, 0.055] 0.9380, [−0.043, 0.033]

<table>
<thead>
<tr>
<th>Processing Methods (Labeled Ratio)</th>
<th>C</th>
<th>g (s)</th>
<th>Normalized Mutual Information (NMI)</th>
<th>Accuracy (ACC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CK-means-ASCASVM (0.5)</td>
<td>31.4351</td>
<td>1.2567</td>
<td>0.8838, [−0.035, 0.061]</td>
<td>0.9333, [−0.056, 0.043]</td>
</tr>
<tr>
<td>SSFCM-ASCASVM (0.5)</td>
<td>73.4361</td>
<td>0.2825</td>
<td>0.8919, [−0.044, 0.071]</td>
<td>0.9386, [−0.044, 0.026]</td>
</tr>
<tr>
<td>FCM-ASCASVDD (0.5)</td>
<td>0.9892</td>
<td>1005.0437</td>
<td>0.6323, [−0.044, 0.071]</td>
<td>0.8421, [−0.053, 0.047]</td>
</tr>
<tr>
<td>CK-means-ASCASVDD (0.5)</td>
<td>0.0303</td>
<td>0.0025</td>
<td>0.8877, [−0.041, 0.048]</td>
<td>0.9366, [−0.044, 0.032]</td>
</tr>
<tr>
<td>SSFCM-ASCASRD-SVDD (0.5)</td>
<td>0.5104</td>
<td>0.1086</td>
<td>0.8601, [−0.083, 0.069]</td>
<td>0.9175, [−0.064, 0.053]</td>
</tr>
<tr>
<td>SSFCM-ASCASRD-SVDD (0.5)</td>
<td>0.2456</td>
<td>18.3967</td>
<td>0.8686, [−0.043, 0.070]</td>
<td>0.9240, [−0.041, 0.047]</td>
</tr>
<tr>
<td>SSFCM-ASCASID-SVDD (0.5)</td>
<td>0.8195</td>
<td>51.0377</td>
<td>0.9102, [−0.020, 0.026]</td>
<td>0.9550, [−0.013, 0.016]</td>
</tr>
<tr>
<td>SSFCM-ASCASID-SVDD (0.5)</td>
<td>0.1523</td>
<td>0.0011</td>
<td>0.9254, [−0.036, 0.021]</td>
<td>0.9649, [−0.029, 0.012]</td>
</tr>
</tbody>
</table>

**Table 5. Fault diagnosis results with different methods.**

**Figure 10.** Comparison of evaluation values for different methods.

**Figure 11.** Boxplots of evaluation values for different methods, the x-axis tick labels correspond to: 1: semi-supervised fuzzy C-means clustering, adaptive sine cosine algorithm and support vector machine (SSFCM-ASCASVM) (0.3); 2: constrained-K-means, adaptive sine cosine algorithm and support vector data description with improved decision strategy (CK-means-ASCASVDD) (0.3); 3: semi-supervised fuzzy C-means clustering, adaptive sine cosine algorithm and support vector data description with relative distance decision strategy (SSFCM-ASCASRD-SVDD) (0.3); 4: semi-supervised fuzzy C-means clustering, sine cosine algorithm and support vector data description with improved decision strategy (SSFCM-ASCASID-SVDD) (0.3); 5: semi-supervised fuzzy C-means clustering, adaptive sine cosine algorithm and support vector data description with improved decision strategy (SSFCM-ASCASID-SVDD) (0.3); 6: constrained-K-means, adaptive sine cosine algorithm and support vector data description with improved decision strategy (CK-means-ASCASVDD) (0.3); 7: SSFCM-ASCASVM (0.5); 8: fuzzy C-means clustering, adaptive sine cosine algorithm and support vector data description with improved decision strategy (FCM-ASCASVDD) (0.5); 9: CK-means-ASCASVDD (0.5); 10: SSFCM-ASCASRD-SVDD (0.5); 11: SSFCM-ASCASRD-SVDD (0.5); 12: SSFCM-ASCASID-SVDD (0.5); 13: SSFCM-ASCASID-SVDD (0.5).
5. Discussion

The superiority of the proposed ASCA and ID-SVDD methods have been effectively demonstrated by the above analysis. Likewise, the accuracy of the diagnosis model has also been verified by contrastive experiments on various fault sizes and locations. This preliminary study mainly deals with the fault diagnosis of vibration signals on the rated load. For the practical application in real industrial operating conditions, variable load conditions should also be considered. Those application situations will be researched in our future work. Besides, the proposed ASCA algorithm can be optimized through better strategies to further improve the convergence speed and global search ability. Moreover, the multi-objective optimization that has been widely utilized in the field of controlling [48–50] could be implemented in fault diagnosis, which is expected to improve the accuracy and reduce the variance of the outputs of the model [51].

6. Conclusions

A novel fault diagnosis method for rotating bearing based on semi-supervised clustering and support vector data description with adaptive parameter optimization and improved decision strategy is presented in this study. Due to the non-stationarity of the original signals, signals collected from different types of faults were firstly split by VMD into sets of IMFs, before which process, the decomposing mode number K was determined by central frequency observation method. Next, the FuzzyEn values of all IMFs were calculated to construct the feature vectors of different types of faults. Subsequently, all training samples were labeled by using SSFCM under the supervision of partly labeled samples. Finally, SVDD with adaptive parameter optimization and improved decision strategy, i.e., ASCA-ID-SVDD, was proposed to diagnose different fault samples. In an engineering application, the contrastive experiments were implemented between the proposed SSFCM-ASCA-ID-SVDD method and other relevant methods. In the sequel, nine types of faults with different locations and sizes were successfully diagnosed, and the results show that the proposed method exhibited the best performance in both NMI and ACC evaluations. Particularly, when training samples with the ratio of 0.5 that are randomly selected as the labeled samples, the NMI value of the proposed method is 0.0335, 0.2913, 0.0377, 0.0568, and 0.0152 higher than that of SSFCM-ASCA-SVM, FCM-ASCA-ID-SVDD, CK-means-ASCA-ID-SVDD, SSFCM-ASCA-RD-SVDD, and SSFCM-SCA-ID-SVDD respectively. Likewise, the ACC value of the proposed method is 0.0263, 0.1228, 0.0263, 0.0409, and 0.0099 higher than that of other methods respectively. Moreover, the boxplots of different methods illustrate the precision and stability of the proposed method. In summary, the proposed method shows better superiority to other methods and achieves preferable diagnostic performance.

Author Contributions: J.T. was responsible for experiments and paper writing, W.F. conceived the research and designed methodology, K.W. helped analyze results, X.X. and W.H. provided recommendations for this research, Y.S. participated in the discussion.

Funding: This work was funded by the National Natural Science Foundation of China (NSFC) (51741907, 51709122), Hubei Provincial Major Project for Technical Innovation (2017AAA132), the Open Fund of Hubei Provincial Key Laboratory for Operation and Control of Cascaded Hydropower Station (2017KJX06), Research Fund for Excellent Dissertation of China Three Gorges University (2019SSPY070).

Conflicts of Interest: The authors declare no conflict of interest.

References
5. Fu, W.; Tan, J.; Li, C.; Zou, Z.; Li, Q.; Chen, T. A hybrid fault diagnosis approach for rotating machinery with the fusion of entropy-based feature extraction and SVM optimized by a chaos quantum sine cosine algorithm. Entropy 2018, 20, 626. [CrossRef]


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).