Research Article

Local Similarity-Based Fuzzy Multiple Kernel One-Class Support Vector Machine

Qiang He ¹, Qingshuo Zhang ², Hengyou Wang ¹ and Changlun Zhang ^{1,3}

¹School of Science, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

²Faculty of Information Technology, Macau University of Science and Technology, Macao 999078, China

³Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture, Beijing 100044, China

Correspondence should be addressed to Qiang He; heqiang@bucea.edu.cn

Received 25 August 2020; Revised 27 September 2020; Accepted 14 October 2020; Published 28 October 2020

Academic Editor: Tongqian Zhang

Copyright © 2020 Qiang He 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.

One-class support vector machine (OCSVM) is one of the most popular algorithms in the one-class classification problem, but it has one obvious disadvantage: it is sensitive to noise. In order to solve this problem, the fuzzy membership degree is introduced into OCSVM, which makes the samples with different importance have different influences on the determination of classification hyperplane and enhances the robustness. In this paper, a new calculation method of membership degree is proposed and introduced into the fuzzy multiple kernel OCSVM (FMKOCSVM). The combined kernel is used to measure the local similarity between samples, and then, the importance of samples is determined based on the local similarity between training samples, so as to determine the membership degree and reduce the impact of noise. The proposed membership requires only positive data in the calculation process, which is consistent with the training set of OCSVM. In this method, the noise has a smaller membership value, which can reduce the negative impact of noise on the classification boundary. Simultaneously, this method of calculating membership has a higher efficiency. The experimental results show that FMKOCSVM based on proposed local similarity membership is efficient and more robust to outliers than the ordinary multiple kernel OCSVMs.

1. Introduction

Anomaly detection is an important aspect of data mining. It is used to find objects in a data set that are significantly different from other data to achieve the purpose of preventing abnormal events. At present, the application of anomaly detection in the field of medicine and biological systems is of great significance, and it has been successfully applied to protein detection, [1] cancer screening, [2] and health monitoring [3]. The essence of anomaly detection is a classification algorithm suitable for processing data with an extremely imbalanced class. Complex biological systems usually have this feature. For example, the data of an infectious disease model may include characteristic data of patients and characteristic data of nonpatients. But, in real life, there are far more healthy people than patients. Timely and effective detection of patients with infectious diseases is

an effective way to prevent the outbreak of infectious diseases.

Support vector machine (SVM) [4, 5] is a classical classification algorithm, but its performance will deteriorate when dealing with a one-class classification problem or distribution imbalance data. Among the solutions to one-class classification problems, there are density estimation-based methods and support vector-based methods. The method based on support vector is popular because of its simplicity and high efficiency. There are two models for this method: (1) one-class support vector machine (OCSVM) [6]; (2) support vector data description (SVDD) [7]. The goal of SVDD is to find a minimum hypersphere that contains all target samples. The main idea of OCSVM is to take the origin of the feature space as a representative of the abnormal data and then separate the target sample from the origin at the maximum margin. This paper focuses on OCSVM.

Like SVM, OCSVM is also sensitive to noise, which is due to the assumption that each sample has the same importance or weight during training. Introducing fuzzy membership in SVM and constructing the fuzzy support vector machine (FSVM) [8] is one of the effective methods to solve this problem. The calculation methods of fuzzy membership are mainly focused on two kinds of classification problems [9–11]. For example, the heuristic function derived from the centered kernel alignment is used to calculate the dependency relationship between the data point and its label to calculate the fuzzy membership [9]. In [10], the membership degree of each sample point is determined by the lower approximation operator of a fuzzy rough set based on Gaussian kernel. As shown in [11], entropy is used to measure the class determinacy of samples. Samples with a higher class determinacy are assigned to a larger fuzzy membership. Generally, in order to improve the robustness of OCSVM, different weights are assigned to training samples, which are called weighted one-class support vector machine (WOCSVM) [12–14]. WOCSVM reduces the impact of noise by assigning lower weights to the noise [12]. As shown in [14], using prior knowledge to assign different weights to the samples, the weight is only related to the distribution knowledge of the neighbors, which is only determined by the k -nearest neighbors of the instance. In recent years, there are many studies on calculating sample weight in one-class classification problems [15–18]. As shown in [19], membership based on fuzzy rough set theory [20, 21] is used as a weight in OCSVM.

The above method improves the robustness of OCSVM to a certain extent but has some limitations. For example, when the amount of data is too large, it [14] is too inefficient when calculating sample weights. The authors in [19] use abnormal data when calculating the membership degree of the sample. In this paper, a novel strategy is proposed to solve the problem of poor robustness of OCSVM; that is, membership degree is introduced into the model. Different from the above membership calculation method, this method only uses one category of data, which fully adapts to the characteristics of OCSVM. The membership calculation method proposed is related to the local density of the data, which is obtained by the local similarity of the training data. We take an S-type function based on local density as a membership function.

OCSVM uses kernel trick to solve the nonlinear separability problem, but it also brings the problem of kernel selection. Multiple kernel learning method [22–26] is used to solve this problem, that is, the multiple kernel one-class support vector machine (MKOCSVM) [27].

The main work of this paper is as follows:

- (1) The multiple kernel learning and membership degree are introduced into OCSVM at the same time, and the fuzzy multiple kernel one-class support vector machine (FMKOCSVM) model is constructed to solve the problem of core selection and noise sensitivity.
- (2) A novel method of membership calculation is proposed, which is based on local similarity.

- (3) We illustrate the effectiveness of this degree of membership in the figure.
- (4) According to the maximum similarity between the combination kernel and the ideal kernel, the weight coefficients of the multiple kernels are determined.
- (5) It is proved by experiments that the method proposed in this paper performs better than the method of fuzzy membership and the model without membership.

The combined kernel more fully characterizes the data than the single kernel. Using multiple kernel functions at the same time can solve the difficult problem of selecting the kernel function and its parameters, and this method can be applied to different sample information.

The rest of this paper is organized as follows: Section 2 introduces the knowledge of OCSVM, MKOCSVM, and FMKOCSVM. The formulation and algorithm of our FMKOCSVM based on local similarity are detailed in Section 3; and Section 4 reports the experimental results, followed by the conclusion in Section 5.

2. Related Information

2.1. One-Class Support Vector Machine. Compared with the SVM, OCSVM is suitable for dealing with the problem of data category imbalance or one-class classification. The main idea is to first map the data from the original space to the feature space through nonlinear mapping and then take the origin of the feature space as the representative of outliers to find an optimal classification hyperplane in the feature space, in which the image of normal data can be separated from the origin by the maximum margin. A graphical illustration is shown in Figure 1. Figure 1(a) shows the description of the classification in the original space. Figure 1(b) shows the description of the classification in the feature space.

Supposed the training samples $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l\} \in \mathcal{R}^n$ (n is the dimension of $\mathbf{x}_i$), where l is the number of training samples. $\phi(\cdot)$ is a function that maps samples to feature spaces. Let ω, ρ denote normal vector and bias term for classification hyperplane in the feature space. The classification hyperplane is expressed as $\omega^T \phi(\mathbf{x}) - \rho = 0$.

The goal is to maximize the distance between the classification hyperplane and origin. Then, OCSVM needs to solve the following convex programming [6]:

$$\begin{aligned} \min_{\omega, \rho, \xi} \quad & \frac{1}{2} \|\omega\|^2 + \frac{1}{\nu l} \sum_{i=1}^l \xi_i - \rho, \\ \text{s.t.} \quad & \omega^T \phi(\mathbf{x}_i) \geq \rho - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, l. \end{aligned} \tag{1}$$

Here, ξ_i is the slack variable, which means that outliers are allowed to exist, and $\nu \in (0, 1]$ is a parameter to control the proportion of support vector and error points. Using the Lagrange multiplier method, the dual problem of the above optimization problem can be written as follows:

FIGURE 1: A graphical illustration of the one-class support vector machine. The black dot represents the target sample, the box represents the abnormal sample, and the black solid line represents the boundary.

$$\begin{aligned} \min_a \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j), \\ \text{s.t.} \quad & \sum_{i=1}^l \alpha_i = 1, 0 \leq \alpha_i \leq \frac{1}{vl}, \quad i = 1, 2, \dots, l. \end{aligned} \quad (2)$$

Because the above process needs to meet the KKT (Karush–Kuhn–Tucker) condition,

$$\begin{cases} \alpha_i \geq 0, \\ \omega^T \phi(\mathbf{x}_i) - \rho + \xi_i \geq 0, \\ \alpha_i [\omega^T \phi(\mathbf{x}_i) - \rho + \xi_i] = 0, \\ \gamma_i \geq 0, \\ \xi_i \geq 0, \\ \gamma_i \xi_i = 0. \end{cases} \quad (3)$$

The solution of equation (3) corresponds to sample $\mathbf{x}_i$, and there is always $\alpha_i = 0$ or $\omega^T \phi(\mathbf{x}_i) = \rho - \xi_i$. When $\alpha_i = 0$, the sample $\mathbf{x}_i$ has no effect on the hyperplane. When $\alpha_i > 0$, $\omega^T \phi(\mathbf{x}_i) = \rho - \xi_i$ must be true. In this case, this sample is called a support vector. If $\alpha_i < 1/vl$, then $\gamma_i > 0$; there must be $\xi_i = 0$; that is, the sample is located on the maximum separation boundary; if $\alpha_i = 1/vl$, then $\gamma_i = 0$; in this case, when $\xi_i > 0$, the sample $\mathbf{x}_i$ is misclassified, which is called the boundary support vector. As shown in Figure 2, when different values are used, the positions of corresponding sample points are different.

Let N_{SV} , N_{BSV} represent the total number of support vectors and the total number of boundary support vectors, respectively. The maximum value of α_i is $1/vl$ and has the following constraints:

$$\sum_{i=1}^l \alpha_i = 1, 0 \leq \alpha_i \leq \frac{1}{vl}, \quad i = 1, 2, \dots, l. \quad (4)$$

So we can have the following inequality:

$$N_{BSV} \frac{1}{vl} \leq \sum_{i=1}^l \alpha_i \leq N_{SV} \frac{1}{vl}, \quad i = 1, 2, \dots, l. \quad (5)$$

Multiplying ν on both sides of equation (5) gives

FIGURE 2: The definitions of support vectors. The black dot represents the target sample, and the black dot with a dotted circle represents support vector.

$$N_{BSV} \frac{1}{l} \leq \nu \leq N_{SV} \frac{1}{l}, \quad i = 1, 2, \dots, l. \quad (6)$$

It can be known from equation (6) that the value of ν determines the lower bound of the total support vector ratio and the upper bound of the boundary support vector ratio:

$$\omega = \sum_{i=1}^l \alpha_i \phi(\mathbf{x}_i). \quad (7)$$

The normal vector of the hyperplane can be obtained by using equation (7).

Let $\mathbf{x}_t^{SV}$ denote the t th support vector located on the maximum spacing plane ($0 < \alpha_t < 1/vl$) and the bias term of the hyperplane obtained according to

$$\rho = \frac{1}{n_{SV}} \sum_{t=1}^{n_{SV}} \omega^T \phi(\mathbf{x}_t^{SV}) = \sum_{t=1}^{n_{SV}} \alpha_t k(\mathbf{x}_t, \mathbf{x}_t^{SV}). \quad (8)$$

Thus, the decision function can be written as

$$\begin{aligned} f(\mathbf{x}) &= \text{sgn}(\boldsymbol{\omega}^T \boldsymbol{\phi}(\mathbf{x}) - \rho) \\ &= \text{sgn}\left[\sum_i^l \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho\right]. \end{aligned} \quad (9)$$

For a given test sample $\mathbf{x}$, after substituting $\mathbf{x}$ into equation (9), when $f(\mathbf{x})$ returns +1, the sample is judged as a normal point; when $f(\mathbf{x})$ returns -1, the sample is judged as an abnormal point.

2.2. Multiple Kernel One-Class Support Vector Machine. MKOCSVM replaces the single kernel function in the conventional OCSVM with a combined kernel, which can effectively avoid the difficulty in selecting the kernel function and its parameters.

The forms of combined kernel function include linear combination and nonlinear combination, [28] expressed as

$$k_\eta(\mathbf{x}, \mathbf{z}) = \sum_{m=1}^P \eta_m k_m(\mathbf{x}, \mathbf{z}), \quad \eta_m \geq 0, \quad (10)$$

$$k_\eta(\mathbf{x}, \mathbf{z}) = \prod_{m=1}^P [k_m(\mathbf{x}^m, \mathbf{z}^m)]^{\eta_m}. \quad (11)$$

Here, η_m is the kernel weight of the m th basic kernel. The MKOCSVM model can be formulated as

$$\begin{aligned} \min_{\boldsymbol{\omega}_\eta, \rho, \xi} \quad & \frac{1}{2} \|\boldsymbol{\omega}_\eta\|^2 + \frac{1}{\nu l} \sum_{i=1}^l \xi_i - \rho, \\ \text{s.t.} \quad & \boldsymbol{\omega}_\eta^T \boldsymbol{\phi}_\eta(\mathbf{x}_i) \geq \rho - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, l. \end{aligned} \quad (12)$$

Here, $k_\eta(\mathbf{x}_i, \mathbf{x}_j) = \langle \boldsymbol{\phi}_\eta(\mathbf{x}_i), \boldsymbol{\phi}_\eta(\mathbf{x}_j) \rangle$. The function $f(\mathbf{x})$ can be written as

$$\begin{aligned} f(\mathbf{x}) &= \text{sgn}(\boldsymbol{\omega}_\eta^T \boldsymbol{\phi}(\mathbf{x}) - \rho) \\ &= \text{sgn}\left[\sum_i^l \alpha_i k_\eta(\mathbf{x}_i, \mathbf{x}) - \rho\right]. \end{aligned} \quad (13)$$

To seek the optimal combination weight for each basis kernel, the authors in [27] suggest optimizing the maximum kernel-target alignment value of the combination kernel and the ideal kernel, that is, solving the following objective function [29]:

$$\begin{aligned} \max_{\eta} A(\mathbf{K}_\eta, \mathbf{y}\mathbf{y}^T) &= \frac{\langle \sum_{m=1}^P \eta_m \mathbf{K}_m, \mathbf{y}\mathbf{y}^T \rangle_F}{\sqrt{\langle \sum_{m=1}^P \eta_m \mathbf{K}_m, \sum_{m=1}^P \eta_m \mathbf{K}_m \rangle_F \langle \mathbf{y}\mathbf{y}^T, \mathbf{y}\mathbf{y}^T \rangle_F}} \\ &= \frac{\sum_{m=1}^P \sum_{i,j=1}^l \eta_m k_{ij}^m}{l \sqrt{\sum_{m,s=1}^P \eta_m \eta_s \langle \mathbf{K}_m, \mathbf{K}_s \rangle_F}} \\ \text{s.t.} \quad & \eta_m \geq 0, \quad m = 1, 2, \dots, P. \end{aligned} \quad (14)$$

Here, $\mathbf{K}_m$ is the kernel matrix. $\langle \mathbf{K}_1, \mathbf{K}_2 \rangle$ is the Frobenius inner product between two matrices, which is given by

$$\langle \mathbf{K}_1, \mathbf{K}_2 \rangle_F = \sum_{i=1}^l \sum_{j=1}^l k_1(\mathbf{x}_i, \mathbf{x}_j) k_2(\mathbf{x}_i, \mathbf{x}_j). \quad (15)$$

Only solution equation (16) is needed to obtain the optimal combination weight:

$$\begin{aligned} \max_{\eta} \quad & \sum_{m=1}^P \sum_{i,j=1}^l \eta_m k_{ij}^m - \sum_{m,s=1}^P \eta_m \eta_s \langle \mathbf{K}_m, \mathbf{K}_s \rangle_F - \delta \sum_{m=1}^P \eta_m^2, \\ \text{s.t.} \quad & \eta_m \geq 0, \quad m = 1, 2, \dots, P. \end{aligned} \quad (16)$$

Here, δ is a regularization coefficient.

2.3. Fuzzy Multiple Kernel One-Class Support Vector Machine.

Let s_i denote membership degree of the sample $\mathbf{x}_i$, then the training set can be expressed as $\{(\mathbf{x}_1, s_1), (\mathbf{x}_2, s_2), \dots, (\mathbf{x}_l, s_l)\} \in \mathfrak{R}^n$, where $s_i \in [0, 1]$. The FMKOC SVM needs to solve the following optimal programming:

$$\begin{aligned} \min_{\boldsymbol{\omega}_\eta, \xi} \quad & \frac{1}{2} \|\boldsymbol{\omega}_\eta\|^2 + \frac{1}{\nu l} \sum_{i=1}^l s_i \xi_i - \rho, \\ \text{s.t.} \quad & \boldsymbol{\omega}_\eta^T \boldsymbol{\phi}_\eta(\mathbf{x}_i) \geq \rho - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, l. \end{aligned} \quad (17)$$

When $s_i = 1, i = 1, 2, \dots, l$, the FMKOC SVM degenerates to the normal MKOCSVM.

After introducing the Lagrange multiplier $\alpha_i \geq 0, \gamma_i \geq 0$, for each inequality constraint, the Lagrange function of equation (17) is

$$L(\omega_\eta, \xi, \rho, \alpha, \gamma) = \frac{1}{2} \|\omega_\eta\|^2 + \frac{1}{\nu l} \sum_{i=1}^l s_i \xi_i - \rho - \sum_{i=1}^l \gamma_i \xi_i - \sum_{i=1}^l \alpha_i [\omega_\eta^T \phi_\eta(\mathbf{x}_i) - \rho + \xi_i]. \quad (18)$$

Setting the derivatives with respect to ω_η , ξ , and ρ to zero, then we can obtain

$$\begin{aligned} \omega_\eta &= \sum_{i=1}^l \alpha_i \phi_\eta(\mathbf{x}_i), \\ \sum_{i=1}^l \alpha_i &= 1, \\ \alpha_i &= \frac{1}{\nu l} s_i - \gamma_i. \end{aligned} \quad (19)$$

Substituting equation (19) into equation (18), the dual form of equation (17) can be written as

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j k_{ij} \\ \text{s.t.} \quad & \sum_{i=1}^l \alpha_i = 1 \\ & 0 \leq \alpha_i \leq \frac{1}{\nu l} s_i, \quad i = 1, 2, \dots, l. \end{aligned} \quad (20)$$

Obviously, the only difference between the dual problems of MKOCSVM and FMKOCSVM is the upper bound of α_i . The upper bound of α_i becomes $s_i/\nu l$ in equation (20). The function $f(\mathbf{x})$ can be written as

$$f(\mathbf{x}) = \text{sgn} \left[\sum_{i=1}^l \alpha_i k_{ij}(\mathbf{x}_i, \mathbf{x}) - \rho \right]. \quad (21)$$

In FMKOCSVM, when the noise has a lower membership during training, the negative effect of noise on the classification hyperplane can be reduced.

3. Training FMKOCSVM with Local Similarity-Based Membership

Noise in the training set may not belong to any class at all. Therefore, if these samples with uncertainty are distributed near the edges of the target data, the model will overfit. To alleviate this phenomenon, this paper assigns membership to each training point, which makes the samples play a different role in training and reduces the negative impact of noise. In this section, we first introduce the calculation method of membership based on local similarity in detail and then propose the FMKOCSVM algorithm using membership based on local similarity.

3.1. Local Similarity-Based Memberships. Assume the target sample $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l\} \in \mathcal{R}^n$. Let $\mathbf{K}_\eta$ represent the multiple kernel matrix defined by $\mathbf{K}_\eta = k_\eta(\mathbf{x}_i, \mathbf{x}_j)_{l \times l}$, where the expression of multiple kernel function is equation (10).

Let all the elements of the upper triangle of the multiple kernel be sorted from large to small, i.e., $\mathbf{K}_\eta = k_\eta(\mathbf{x}_i, \mathbf{x}_j)_{i < j}$, and then write it as a vector $G = \{g_1, g_2, \dots, g_t\}$, $t = l(l-1)/2$.

Next, define a constant $\varepsilon \in [0, 1]$ and let $h = [\varepsilon t]$, where $[\varepsilon t]$ is the integral part of εt . $\theta = g_h$ is a threshold. For each sample $\mathbf{x}$, let u represent the total number of $k_{ij} \geq \theta$, $j = 1, 2, \dots, l$. In other words, u_i represents the number of samples in the target sample whose similarity with the sample $\mathbf{x}_i$ is greater than or equal to the threshold.

The kernel $k_\eta(\mathbf{x}_i, \mathbf{x}_j)$ measures the similarity between two target samples $\mathbf{x}_i$ and $\mathbf{x}_j$, and a large kernel indicates a large similarity. If the membership degree of the sample $\mathbf{x}_i$ to the target class is higher, it is obvious that more samples are similar to the sample $\mathbf{x}_i$ in the input sample, i.e., the greater the value of u_i . In other words, a sample with a higher value of u should have a greater contribution to the classification boundary, the penalty for misclassification of such samples is greater, and the noise will have a smaller value of u .

Therefore, we take u as a measure function, which measures the importance of the target sample to the classification hyperplane. Obviously, the value of u cannot be directly used as a membership degree of FMKOCSVM. We use an S-type function to map this measure into the membership degree in the unit interval. At the same time, this S-type function increases the difference between membership degrees of samples with different importance. The membership function is written as

$$s_i = \frac{1}{1 + \exp(-\tau((u_i/U) - (1/2)))}, \quad (22)$$

where $U = \max\{u_i\}$. $\tau > 0$ is a constant. The value range is $(0, 1)$. Figure 3 describes the distribution of membership values by τ taking different values. According to Figure 3, when $\tau = 10$, the distribution of membership value is the best. Algorithm 1 lists the detailed calculation process of membership based on local similarity.

Since noise has a low degree of membership to the target class, there are few similar instances in the input data. In other words, the noise will get a smaller membership value. Therefore, we proposed a method that can make the noise have less influence on the classification boundary. More importantly, because the training data of OCSVM only include target samples, the traditional method of calculating membership degree is not suitable for OCSVM. However, our method of membership based on local similarity only uses the features of the target data and does not involve the information of class, which is very suitable for a one-class classification problem. Furthermore, the proposed method has obvious high efficiency.

We analyze the computational complexity of Algorithm 1 with the O notation. First, the computational

FIGURE 3: The distribution of membership with different values of τ . The horizontal axis is the values of u/U . The vertical axis is the corresponding membership value. The black dashed line is the distribution of membership when $\tau = 1$. The red solid line is the distribution of membership when $\tau = 10$. The blue dash-dotted line is the distribution of membership when $\tau = 20$. The cyan dotted line is the distribution of membership degree when $\tau = 30$.

Input: the training set $\{x_1, x_2, \dots, x_l\}$, the kernel function set $\{k_1, k_2, \dots, k_p\}$,
The kernel weight $\eta_m, m = 1, 2, \dots, P, \varepsilon \in [0, 1]$
Output: the membership vector $s_i, i = 1, 2, \dots, l$

- (1) Preprocess the training set
- (2) Calculate the combined kernel matrix $K_\eta = (k_{\eta_{ij}})_{l \times l}$ according to equation (10)
- (3) Sort $K_\eta = (k_{\eta_{ij}})_{i < j}$ from large to small as $G = \{g_1, g_2, \dots, g_t\}, t = l(l-1)/2$
- (4) Calculate the constant h according to $h = \lceil \varepsilon t \rceil$, and fixed threshold according to $\theta = g_h$
- (5) for $i = 1 : l$ do
- (6) initialize $u_i = 0$
- (7) for $j = 1 : l$ do
- (8) if $k_{ij} \geq g_h$ then
- (9) $u_i = u_i + 1$
- (10) end if
- (11) end for
- (12) end for
- (13) Calculate $U = \max_i \{u_i\}$
- (14) for $i = 1 : l$ do
- (15) Calculate the membership degree s_i of the sample x_i according to equation (22)
- (16) end for
- (17) end

ALGORITHM 1: Local similarity-based memberships.

complexity of calculating the multiple kernel matrix in step 2 is $O(l^2)$. Second, the average computational complexity of multiple kernel matrix sorting in step 3 is $O(t^2)$. Third, it costs $O(l^2)$ to calculate u in steps 4 to 12. Finally, calculating the memberships in steps 13 to 16 costs $O(l)$. Hence, the total computational complexity of local similarity-based membership degree is

$$\begin{aligned} O(l^2) + O(t^2) + O(l^2) + O(l) &= O(l^2) + O(l^4) + O(l^2) + O(l) \\ &= O(l^4). \end{aligned} \quad (23)$$

Compared with classification performance, $O(l^4)$ is also acceptable.

Input: the training set $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l\}$, the kernel function set $\{k_1, k_2, \dots, k_p\}$, the kernel parameter set $\{\sigma_1, \sigma_2, \dots, \sigma_p\}$, the test set T

Output: the classification results of T : y

- (1) Preprocess the training set
- (2) for $m = 1 : P$ do
- (3) Calculate the corresponding kernel matrix $\mathbf{K}_m$ of σ_m
- (4) end for
- (5) Substituting the kernel matrix $\mathbf{K}_m, m = 1, 2, \dots, P$ into equation (16)
- (6) Calculate the kernel weight vector $\eta_m, m = 1, 2, \dots, P$ according to equation (16)
- (7) Calculate the combined kernel matrix $\mathbf{K}_\eta$ according to equation (10)
- (8) for each $x_i \in \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l\}$ do
- (9) Calculate the membership degree s_i of the sample x_i according to equation (22)
- (10) end for
- (11) Train the FMKOC SVM with the fuzzy membership values.
- (12) Calculate the classification results of test set according to equation (21)
- (13) end

ALGORITHM 2: FMKOC SVM_LS.

FIGURE 4: The classification performance of MKOC SVM and FMKOC SVM_LS without noises. The magenta solid line is the boundary. The blue dots are the target samples.

3.2. Overall Procedure of FMKOC SVM Based on Local Similarity. The detailed process of FMKOC SVM based on membership degree of local similarity is listed in Algorithm 2. In the following part of this paper, we use FMKOC SVM_LS to represent the proposed algorithm.

In Figure 4, the classification performance of MKOC SVM is shown in Figure 4(a) and the classification performance of FMKOC SVM_LS proposed by us is shown in Figure 4(b). The combined kernel function is composed of seven Gaussian kernels with a width of $\{2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0, 2^1\}$. Parameter ν is set to 0.02. The value of the regularization coefficient δ in equation (16) is set to 100. In order to reduce the cost of tuning extraparameter, we set ε to 0.2 and τ to 10 directly. Obviously, FMKOC SVM_LS has a tighter boundary than MKOC SVM. In Figure 4(b), some outliers are identified by FMKOC SVM_LS. However, MKOC SVM does not identify any outliers, and there are a lot of gaps inside the boundary.

After adding 10% Gaussian noise to the training set, the results are shown in Figure 5.

The parameter settings in Figure 5 are the same as in Figure 4. As we can see, when there are noises in the training set, the classification ability of FMKOC SVM_LS is much better than that of MKOC SVM. In Figure 5(a), MKOC SVM distinguishes all noises as target data, which makes the performance of MKOC SVM very bad. In Figure 5(b), we can see that most of the noises have a very small membership value, and the negative effect of noise on the boundary is weak. Therefore, FMKOC SVM_LS can improve the robustness of MKOC SVM.

In the next section, we will further prove through experiments that our proposed method of calculating membership is better than the previous method.

4. Experiments

4.1. Experiment Setup

4.1.1. Approaches. We compared FMKOC SVM_LS with the following methods: (1) MKOC SVM: the ordinary multiple kernel one-class support vector machine [27]; (2)

FIGURE 5: The classification performance of MKOCSVM and FMKOCSVM_LS. 10% of Gaussian noise is added. The magenta solid line is the boundary. The blue dots are the target samples. The black crosses are the noises with Gaussian distribution. The number at bottom of the noise represents its membership.

WMKOCSVM: the weighted one-class support vector machine is formed by WOCSVM [14] combined with the multiple kernel function; (3) FMKOCSVM: the fuzzy multiple kernel one-class support vector machine, in which membership is calculated based on a rough set [19]. Because two classes of samples are needed to calculate membership, the training set contains negative class samples. These negative samples are only used to calculate membership.

In MKOCSVM, the parameter ν is determined by 10-fold cross validation, and the value range is $\{0.01, 0.04, 0.07, 0.10, 0.13, 0.16, 0.19\}$. The basic kernel of the multiple kernel function is seven Gauss kernel functions with kernel width $\{2^{-6}, 2^{-5}, 2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 2^0\}$. The parameter δ in the multiple kernel learning algorithm is set to 100. WMKOCSVM, FMKOCSVM, and FMKOCSVM_LS also use these parameters during training. The number of nearest neighbors in WMKOCSVM is set to 10, which is the same as in [14]. In order to avoid increasing the time due to the adjustment of parameters, when calculating the membership based on local similarity in FMKOCSVM_LS, we directly set $\varepsilon = 0.5$ and $\tau = 10$.

4.1.2. Metrics. In this paper, the performance of different approaches is evaluated by three popular metrics, namely, g-mean, AUC, and training time. According to the confusion matrix in Table 1, we can get the true positive rate (TPR) and the false positive rate (FPR). In one-class classification problems, using g-mean and AUC as measures is more accurate than using accuracy:

$$\begin{aligned} \text{TPR} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{FPR} &= \frac{\text{FP}}{\text{FP} + \text{TN}}, \\ g\text{-mean} &= \sqrt{a^+ \times a^-}, \\ \text{AUC} &= \frac{1 + \text{TPR} - \text{FPR}}{2}. \end{aligned} \quad (24)$$

TABLE 1: Confusion matrix.

	Predicted positive	Predicted negative
Positive	TP	FN
Negative	FP	TN

4.1.3. Data Sets. In this section, we selected 14 benchmark data sets, 13 of which are from the UCI machine learning repository. There are three experiments on biological systems. The Heart data set is a data set used for heart disease diagnosis. The Breast data set is a data set used to diagnose whether a patient's breast cancer is benign or malignant. The Biomed data set is used to screen whether it is a carrier. Creditcard_cut is a part of the data set of creditcard fraud detection on Kaggle. Because the original Creditcard data set is too large, we only randomly selected 729 transaction data (483 normal transactions and 249 fraudulent transactions) for the experiment. Table 2 lists the details of these data sets.

For each data set, we use 70% of the positive data as the training set. Then, we randomly selected a part of negative data as the noise in the training set, and the proportion of noise was 10%. The rest of the data is used as the test set. The training set is normalized before training. And the test set is processed according to the standard of the training set.

4.2. Results. In order to obtain stable results, each method has done 20 independent experiments on each data set. The result used in the comparison is the average of the 20 results. Table 3 shows the optimal value of ν obtained through the 10-fold cross validation. In order to get the best results of the four algorithms, ν in Table 3 is used in each experiment.

Table 4 is a comparison of g-mean. Table 5 shows a comparison of AUC values. Table 6 shows the average training time of MKOCSVM, WMKOCSVM, FMKOCSVM, and FMKOCSVM_LS on each data set. Figure 6 is the total training time of 14 data sets of each method.

TABLE 2: Data information.

Number	Data set	Positive samples	Negative samples	Samples	Features
1	Australia	383	307	690	14
2	Balancescaleleft	288	337	625	4
3	Biomed	127	67	194	5
4	Glass	70	144	214	9
5	Heart	160	137	297	13
6	Vowel	48	480	528	10
7	Wine	48	130	178	13
8	Creditcard_cut	483	249	729	30
9	Japan	357	294	651	15
10	Iris	50	100	150	4
11	Breast	458	241	699	9
12	Wdbc	357	212	569	30
13	Pima	500	268	768	8
14	Waveform	300	600	900	21

TABLE 3: The optimal value of ν .

Number	Data set	MKOCSVM	WMKOCSVM	FMKOCSVM	FMKOCSVM_LS
1	Australia	0.19	0.1	0.19	0.07
2	Balancescaleleft	0.19	0.1	0.07	0.04
3	Biomed	0.16	0.04	0.07	0.01
4	Glass	0.07	0.07	0.04	0.01
5	Heart	0.19	0.07	0.19	0.04
6	Vowel	0.01	0.01	0.01	0.01
7	Wine	0.01	0.01	0.01	0.01
8	Creditcard_cut	0.13	0.04	0.07	0.01
9	Japan	0.19	0.1	0.19	0.13
10	Iris	0.01	0.01	0.01	0.01
11	Breast	0.04	0.01	0.04	0.01
12	Wdbc	0.07	0.01	0.04	0.01
13	Pima	0.19	0.1	0.1	0.07
14	Waveform	0.16	0.04	0.07	0.01

TABLE 4: Comparison of g-mean.

Number	Data set	g -mean (mean $\pm$ std (%))			
		MKOCSVM	WMKOCSVM	FMKOCSVM	FMKOCSVM_LS
1	Australia	59.19 $\pm$ 2.22	65.08 $\pm$ 2.39	74.18 $\pm$ 2.68	74.74 $\pm$ 2.51
2	Balancescaleleft	61.38 $\pm$ 2.05	66.17 $\pm$ 2.29	66.73 $\pm$ 3.22	69.99 $\pm$ 3.09
3	Biomed	72.43 $\pm$ 5.59	76.80 $\pm$ 4.53	78.71 $\pm$ 2.98	81.07 $\pm$ 4.15
4	Glass	67.74 $\pm$ 5.41	74.02 $\pm$ 4.38	70.20 $\pm$ 5.34	72.51 $\pm$ 6.64
5	Heart	63.90 $\pm$ 1.79	63.76 $\pm$ 3.90	67.82 $\pm$ 5.05	70.51 $\pm$ 3.45
6	Vowel	79.46 $\pm$ 8.26	79.02 $\pm$ 4.79	80.42 $\pm$ 5.72	84.88 $\pm$ 6.24
7	Wine	81.49 $\pm$ 7.59	81.20 $\pm$ 7.17	82.87 $\pm$ 6.71	85.68 $\pm$ 9.07
8	Creditcard_cut	72.00 $\pm$ 3.37	74.26 $\pm$ 3.90	81.36 $\pm$ 2.18	83.66 $\pm$ 1.91
9	Japan	58.23 $\pm$ 4.08	64.59 $\pm$ 2.77	73.10 $\pm$ 2.91	68.87 $\pm$ 3.52
10	Iris	63.87 $\pm$ 15.23	69.67 $\pm$ 6.95	75.48 $\pm$ 13.65	93.87 $\pm$ 4.76
11	Breast	60.71 $\pm$ 4.85	65.72 $\pm$ 5.28	81.22 $\pm$ 4.54	92.00 $\pm$ 1.93
12	Wdbc	59.06 $\pm$ 5.03	55.31 $\pm$ 6.78	73.08 $\pm$ 4.24	86.41 $\pm$ 1.83
13	Pima	51.92 $\pm$ 1.91	61.03 $\pm$ 2.18	62.39 $\pm$ 2.18	64.92 $\pm$ 2.09
14	Waveform	68.99 $\pm$ 4.39	67.65 $\pm$ 3.49	74.13 $\pm$ 2.52	78.23 $\pm$ 1.87

From Tables 4 and 5, we can find that the performance of FMKOCSVM_LS is the best among the four algorithms, which proves that our membership method can improve the robustness of MKOCSVM. More importantly, WMKOCSVM and FMKOCSVM have only one best result

on 14 data sets, respectively. However, our proposed method has twelve optimal performances.

On the Iris, Breast, and Wdbc data sets, FMKOCSVM_LS shows great advantages. Its g-mean is 27%~32% higher than that of MKOCSVM, 10%~18% higher

TABLE 5: Comparison of AUC value.

Number	Data set	AUC (mean $\pm$ std (%))			
		MKOC SVM	WMKOC SVM	FMKOC SVM	FMKOC SVM_LS
1	Australia	62.25 $\pm$ 1.99	65.43 $\pm$ 2.32	74.35 $\pm$ 2.62	74.88 $\pm$ 2.49
2	Balancescaleleft	64.21 $\pm$ 1.91	66.88 $\pm$ 2.27	68.10 $\pm$ 3.17	70.64 $\pm$ 2.84
3	Biomed	73.27 $\pm$ 5.17	77.10 $\pm$ 4.41	79.05 $\pm$ 2.80	81.22 $\pm$ 4.11
4	Glass	71.26 $\pm$ 4.31	74.29 $\pm$ 4.13	71.26 $\pm$ 5.33	73.63 $\pm$ 6.37
5	Heart	64.81 $\pm$ 1.66	64.08 $\pm$ 3.89	68.00 $\pm$ 5.10	71.09 $\pm$ 2.99
6	Vowel	80.17 $\pm$ 7.32	79.49 $\pm$ 4.65	80.78 $\pm$ 5.64	85.32 $\pm$ 5.71
7	Wine	82.23 $\pm$ 6.84	81.74 $\pm$ 7.02	83.62 $\pm$ 1.95	87.04 $\pm$ 7.17
8	Creditcard_cut	73.86 $\pm$ 2.67	75.25 $\pm$ 3.18	81.44 $\pm$ 2.20	83.70 $\pm$ 1.90
9	Japan	60.84 $\pm$ 3.44	64.86 $\pm$ 2.79	73.24 $\pm$ 2.84	69.03 $\pm$ 3.48
10	Iris	69.90 $\pm$ 10.54	71.75 $\pm$ 6.32	78.24 $\pm$ 9.82	94.17 $\pm$ 4.44
11	Breast	67.87 $\pm$ 2.95	71.04 $\pm$ 3.61	82.53 $\pm$ 3.76	92.34 $\pm$ 1.76
12	Wdbc	65.65 $\pm$ 3.25	63.44 $\pm$ 4.02	75.17 $\pm$ 3.44	86.50 $\pm$ 1.73
13	Pima	57.33 $\pm$ 1.84	61.82 $\pm$ 2.07	62.75 $\pm$ 2.03	65.09 $\pm$ 2.18
14	Waveform	70.70 $\pm$ 3.86	69.80 $\pm$ 2.94	74.35 $\pm$ 2.53	78.80 $\pm$ 1.82

TABLE 6: Average training time.

Number	Data set	Training time (s)			
		MKOC SVM	WMKOC SVM	FMKOC SVM	FMKOC SVM_LS
1	Australia	20.60	48.65	22.67	28.22
2	Balancescaleleft	13.13	30.30	16.85	19.34
3	Biomed	5.43	12.75	6.16	7.68
4	Glass	1.52	5.31	3.17	2.92
5	Heart	7.75	17.87	7.96	9.64
6	Vowel	3.57	5.11	4.58	2.95
7	Wine	1.31	4.89	1.44	3.21
8	Creditcard_cut	38.51	75.93	47.19	54.96
9	Japan	18.81	44.18	20.30	24.42
10	Iris	1.22	4.22	1.70	1.88
11	Breast	32.86	51.27	33.24	47.58
12	Wdbc	18.71	45.89	23.86	28.26
13	Pima	38.49	68.94	51.42	57.08
14	Waveform	14.91	37.20	20.84	21.97

FIGURE 6: The total training time of 14 data sets of MKOC SVM, WMKOC SVM, FMKOC SVM, and FMKOC SVM_LS.

than that of FMKOC SVM, and 23%~31% higher than that of WMKOC SVM. In the corresponding data set, the AUC value of FMKOC SVM_LS also increased significantly. On the Japan data set, although the g-mean of FMKOC SVM_LS is lower than the g-mean of FMKOC SVM, it is still 10% higher than the g-mean of MKOC SVM and 4% higher than the g-mean of WMKOC SVM. AUC value is the same as g-mean. On the Glass data set, WMKOC SVM has the best result, and its result is only about 2% higher than that of FMKOC SVM_LS. However, on the Glass data set, the result of FMKOC SVM_LS is 5% higher than that of MKOC SVM, which proves that our membership calculation method can reduce the impact of noise on classification ability. In the remaining 9 data, the results of FMKOC SVM_LS are the best and have obvious advantages. For example, on the Waveform data set, the g-mean of FMKOC SVM_LS is 10% higher than that of WMKOC SVM and 4% higher than that of FMKOC SVM.

In terms of training time, although our method is not the fastest, FMKOC SVM_LS is still faster than WMKOC SVM. The training time of WMKOC SVM is 1.5 times of that of FMKOC SVM_LS on average. Compared with the training time of MKOC SVM, the increased training time of FMKOC SVM_LS is within the acceptable range.

All of the above proves that the MKOC SVM with membership is more robust. Moreover, our proposed membership based on local similarity is the best.

5. Conclusions

In order to solve the problem of poor robustness of MKOC SVM, this paper proposes a fuzzy multiple kernel one-class support vector machine based on local similarity, in which membership is based on the local similarity of the training data. Firstly, the similarity between samples is measured by combining the kernel matrix. Then, according to the selected threshold, the local similarity of each sample is determined. Finally, an S-type function is used to map the local similarity to the unit interval, and the function value is taken as the membership value. Experiments show that the membership method proposed in this paper can improve the robustness of MKOC SVM. Moreover, compared with the other two methods, our method is optimal.

The difficulty in fuzzy multiple kernel one-class support vector machine lies in how to determine the effective membership. Compared with the previous membership calculation method, only the target data are needed to calculate the membership based on the local similarity of the data, which is consistent with the OCSVM training set. In this method, the noise or outliers are assigned a small membership value, which makes the noise have the weakest impact on the classification boundary. Therefore, the membership method in this paper helps improve the robustness of MKOC SVM. In the next step, we will research the optimization method of parameters in the process of membership calculation based on local similarity.

Data Availability

The data underlying the study can be available upon request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (nos. 62072024 and 61473111), Projects of Beijing Advanced Innovation Center for Future Urban Design, Beijing University of Civil Engineering and Architecture (nos. UDC2019033324 and UDC2017033322), Scientific Research Foundation of Beijing University of Civil Engineering and Architecture (no. KYJJ2017017), Natural Science Foundation of Guangdong Province (no. 2018A0303130026), and Natural Science Foundation of Hebei Province (no. F2018201096).

References

- [1] M. Torrisi, G. Pollastri, and Q. Le, "Deep learning methods in protein structure prediction," *Computational and Structural Biotechnology Journal*, vol. 18, pp. 1301–1310, 2020.
- [2] A. Asuntha and A. Srinivasan, "Deep learning for lung cancer detection and classification," *Multimed Tools Application*, vol. 79, no. 11–12, pp. 1–32, 2020.
- [3] S. Kiranyaz, M. Zabihi, A. B. Rad, T. Ince, R. Hamila, and M. Gabbouj, "Real-time phonocardiogram anomaly detection by adaptive 1D convolutional neural networks," *Neurocomputing*, vol. 411, pp. 291–301, 2020.
- [4] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, pp. 144–152, Pittsburgh, PA, USA, July 1992.
- [5] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, Berlin, Germany, 2nd edition, 1999.
- [6] B. Schölkopf, R. C. Williamson, A. J. Smola et al., "Support vector method for novelty detection," in *Proceedings of the Advances in Neural Information Processing Systems in Advances 12 (NIPS)*, pp. 582–588, Denver, CO, USA, 2000.
- [7] 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.
- [8] C. Lin and S. Wang, "Fuzzy support vector machines," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 464–471, 2002.
- [9] T. Wang, Y. Qiu, and J. Hua, "Centered kernel alignment inspired fuzzy support vector machine," *Fuzzy Sets and Systems*, vol. 394, pp. 110–123, 2020.
- [10] Q. He and C. Wu, "Membership evaluation and feature selection for fuzzy support vector machine based on fuzzy rough sets," *Soft Computing*, vol. 15, no. 6, pp. 1105–1114, 2011.
- [11] Q. Fan, Z. Wang, D. Li, D. Gao, and H. Zha, "Entropy-based fuzzy support vector machine for imbalanced datasets," *Knowledge-Based Systems*, vol. 115, pp. 87–99, 2017.
- [12] M. Bicego and M. A. T. Figueiredo, "Soft clustering using weighted one-class support vector machines," *Pattern Recognition*, vol. 42, no. 1, pp. 27–32, 2009.

- [13] B. Cyganek, "Image Segmentation with a hybrid ensemble of one-class support vector machines," in *Proceedings of the International Conference on Hybrid Artificial Intelligence Systems (HAIS)*, pp. 256–263, San Sebastian, Spain, 2010.
- [14] F. Zhu, J. Yang, C. Gao, S. Xu, N. Ye, and T. Yin, "A weighted one-class support vector machine," *Neurocomputing*, vol. 189, pp. 1–10, 2016.
- [15] M. Cha, J. S. Kim, and J. K. Baek, "Density weighted support vector data description," *Expert Systems with Applications*, vol. 41, no. 7, pp. 3343–3350, 2014.
- [16] J. Yang, T. Deng, and R. Sui, "An adaptive weighted one-class SVM for robust outlier detection," in *Proceedings of the 2015 Chinese Intelligent Systems Conference*, pp. 475–484, Yangzhou, China, 2016.
- [17] T. Hou, Y. Liu, K. Wang, J. Xie, and F. Liu, "A new weighted SVDD algorithm for outlier detection," in *Proceedings of the 28th Chinese Control and Decision Conference*, pp. 5456–5461, Yinchuan, China, 2016.
- [18] S. Yin, X. Zhu, and C. Jing, "Fault detection based on a robust one class support vector machine," *Neurocomputing*, vol. 145, pp. 263–268, 2014.
- [19] Q. Zhang, Q. He, C. Zhang, and H. Wang, "Multiple kernel one-class support vector machine," *Journal of Beijing University of Civil Engineering and Architecture*, vol. 36, no. 1, pp. 82–90, 2020, in Chinese.
- [20] D. Dubois, H. Prade, and R. SŁOWIŃSKI, *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*, Springer, Berlin, Germany, 1992.
- [21] D. Chen, S. Kwong, Q. He, and H. Wang, "Geometrical interpretation and applications of membership functions with fuzzy rough sets," *Fuzzy Sets and Systems*, vol. 193, pp. 122–135, 2012.
- [22] S. K. Siahroudi, P. Z. Moodi, and H. Beigy, "Detection of evolving concepts in non-stationary data streams: A multiple kernel learning approach," *Expert Systems with Applications*, vol. 91, pp. 187–197, 2018.
- [23] Y. G. Zhao, Z. Song, F. Zheng, and L. Shao, "Learning a multiple kernel similarity metric for Kinship verification," *Information Sciences*, vol. 430–431, pp. 247–260, 2018.
- [24] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.
- [25] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the SMO algorithm," in *Proceedings of the International Conference on Machine Learning*, pp. 41–48, New York, NY, USA, 2004.
- [26] J. Qi, X. Liang, and R. Xu, "A multiple kernel learning model based on p -norm," *Computational Intelligence and Neuroscience*, vol. 2018, pp. 1–7, Article ID 1018789, 2018.
- [27] Q. He, Q. Zhang, and H. Wang, "Kernel-target alignment based multiple kernel one-class support vector machine," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 2083–2088, Bari, Italy, 2019.
- [28] H.-Q. Wang, F.-C. Sun, Y.-N. Cai, N. Chen, and L.-G. Ding, "Multiple kernel learning methods," *Acta Automatica Sinica*, vol. 36, no. 8, pp. 1037–1050, 2010, in Chinese.
- [29] L. Chen, D. Chen, and H. Wang, "Alignment based kernel selection for multi-label learning," *Neural Processing Letters*, vol. 49, no. 3, pp. 1157–1177, 2019.