

Research Article

Research on Multiple Classification Based on Improved SVM Algorithm for Balanced Binary Decision Tree

Wenhao Xie ¹, Yanhong She ¹ and Qiao Guo ²

¹School of Science, Xi'an Shiyou University, Xi'an 710065, China

²College of Petroleum Engineering, Xi'an Shiyou University, Xi'an, 710065, China

Correspondence should be addressed to Wenhao Xie; xwhaoxwhao@163.com and Yanhong She; yanhongshe@xsyu.edu.cn

Received 8 January 2021; Accepted 22 March 2021; Published 5 April 2021

Academic Editor: Michele Risi

Copyright © 2021 Wenhao Xie 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.

Support vector machines (SVMs) are designed to solve the binary classification problems at the beginning, but in the real world, there are a lot of multiclassification cases. The multiclassification methods based on SVM are mainly divided into the direct methods and the indirect methods, in which the indirect methods, which consist of multiple binary classifiers integrated in accordance with certain rules to form the multiclassification model, are the most commonly used multiclassification methods at present. In this paper, an improved multiclassification algorithm based on the balanced binary decision tree is proposed, which is called the IBDT-SVM algorithm. In this algorithm, it considers not only the influence of “between-classes distance” and “class variance” in traditional measures of between-classes separability but also takes “between-classes variance” into consideration and proposes a new improved “between-classes separability measure.” Based on the new “between-classes separability measure,” it finds out the two classes with the largest between-classes separability measure and uses them as the positive and negative samples to train and learn the classifier. After that, according to the principle of the class-grouping-by-majority, the remaining classes are close to these two classes and merged into the positive samples and the negative samples to train SVM classifier again. For the samples with uneven distribution or sparse distribution, this method can avoid the error caused by the shortest canter distance classification method and overcome the “error accumulation” problem existing in traditional binary decision tree to the greatest extent so as to obtain a better classifier. According to the above algorithm, each layer node of the decision tree is traversed until the output classification result is a single-class label. The experimental results show that the IBDT-SVM algorithm proposed in this paper can achieve better classification accuracy and effectiveness for multiple classification problems.

1. The Introduction of Support Vector Classifier

Setting a set, which includes m samples: $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$, in which $x_i \in R^n$, $y_i \in (-1, +1)$ ($i = 1, 2, \dots, l$). If the hyperplane $w \cdot x + b = 0$ can separate the samples of this set exactly and the distance from the nearest sample to the hyperplane is the maximum, this hyperplane is called the optimal hyperplane, which is also called the maximum margin hyperplane (see Figure 1).

Different from other classification algorithms, SVM can well solve the problems of data classification in high-dimensional space by the kernel functions. It maps the samples from the input space to the higher dimensional space by this mapping: $\varphi: R_n \rightarrow F$, and it can construct the optimal

classification hyperplane to separate the samples in the high-dimensional feature space. In this way, the problem above is transformed into the following optimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i, \\ \text{s.t.} \quad & y_i (w \cdot \varphi(x_i) + b) - 1 + \xi_i \geq 0, \\ & \xi_i \geq 0, \quad i = 1, \dots, l, \end{aligned} \quad (1)$$

in which C is the penalty parameter, which controls the degree of penalty for the misclassification samples. In addition, the greater the value of C , the greater the penalty for error.

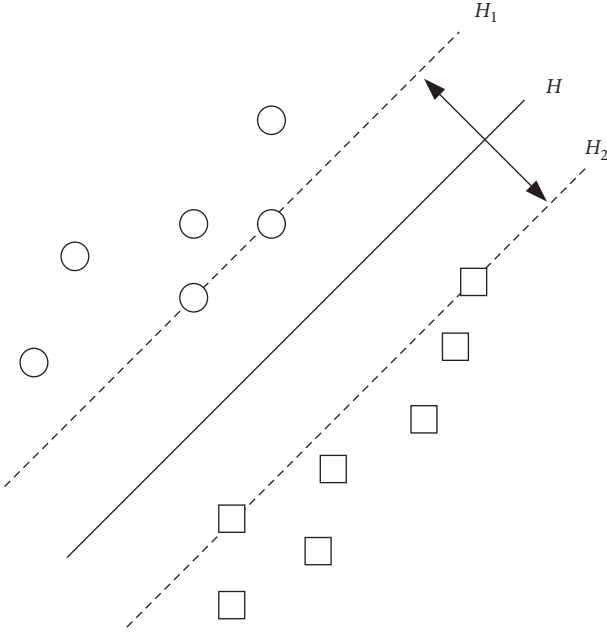


FIGURE 1: The maximum margin hyperplane [1] (reproduced from Xie et al. 2019, under the Creative Commons Attribution License/public domain).

It converts the above problems into the following dual problem:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j K(x_i \cdot x_j) - \sum_{i=1}^l \alpha_i, \\ \text{s.t.} \quad & \sum_{i=1}^l \alpha_i y_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, \dots, l, \end{aligned} \quad (2)$$

in which α_i is the Lagrange multiplier and $K(x_i \cdot x_j) = \varphi(x_i) \cdot \varphi(x_j) \in R$ is the kernel function, which represents the inner product of $\varphi(x_i)$ and $\varphi(x_j)$. If $K(x_i \cdot x_j)$ is positive definite or semidefinite, formula (2) is a convex quadratic programming problem. And the final decision function is as follows:

$$f_i = f(x_i) = \text{sgn} \left(\sum_{j=1}^l \alpha_j y_j K(x_j, x_i) + b \right). \quad (3)$$

2. The Analysis of the Multiclassification SVM Algorithms

2.1. An Overview of the Multiclassification SVM Methods. Currently, there are two main multiclassification methods for SVM including the direct method and the indirect methods [2, 3]. The principles of these methods are briefly introduced as follows:

- (1) The direct method directly modifies the quadratic programming form of SVM, designs the multiple

objective functions, solves a large quadratic programming problem based on the SVM model, and completes the multiple classification problems at one time.

Let $T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (X \times Y)^l$ be the training set, which meets the conditions: $x_i \in X = R^n$, $y_i \in Y = \{1, 2, \dots, M\}$, $i = 1, 2, \dots, l$. It will look for a decision function: $f(x): X = R^n \rightarrow Y$ and implements the mapping from X to Y when $M > 2$.

By modifying the quadratic programming equation, it can obtain the objective function of the optimization problem based on SVM multiclassification as follows [4]:

$$\begin{aligned} \min \quad & \frac{1}{2} \sum_{r=1}^M \|w_r\|^2 + \frac{C}{m} \sum_{i=1}^m \sum_{r \neq y_i} \xi_i^r, \\ \text{s.t.} \quad & (w_{y_i} \cdot x_i) + b_{y_i} \geq (w_r \cdot x_i) + b_r + 2 - \xi_i^r, \\ & \xi_i^r \geq 0, \end{aligned} \quad (4)$$

in which $m \in \{1, \dots, M\} \setminus \{y_i\}$, $y_i \in \{1, \dots, M\}$ is the multiclassification label corresponding to the sample x_i .

This algorithm makes a direct improvement on the original SVM classification model, and the theory is simple. However, due to the large number of variables, the solution process of the objective function is difficult, especially in the case of a large number of samples, which will greatly increase the difficulty of calculation and solution and increase the training time. At the same time, compared with other algorithms, this method does not have a good classification accuracy, so it is not widely used in practice.

- (2) By contrast, the indirect methods design multiple binary classifiers first and then sort and combine the samples according to some rules so as to realize the multilabel identification for them [5]. The advantage of the indirect methods is obvious. The principle of the methods is to integrate multiple binary classifiers, each classification is equivalent to training one binary classifier separately, and processing number of the support vectors and training time are greatly reduced in every time, so it has strong operability and practicability. In recent years, the indirect methods which use the integrated binary classifiers have become the research hotspots of the SVM multiclassification research.

2.2. The Introduction for the Binary Decision Tree Method Based on SVM. There are many indirect classification methods based on SVM: one-versus-one method, that is, OVO-SVM method [6]; one-versus-all method, that is, OVA-SVM [7, 8]; directed acyclic graph method, that is, DAG-SVM method [9, 10], error-correcting output codes method [11, 12], binary decision tree method [13, 14], and

hierarchical multiclass SVM algorithm [15, 16], in which the binary decision tree method is divided into the balanced binary decision tree and the imbalanced binary decision tree.

The principle of the SVM multiclass classification method based on Decision Tree (DT) is to construct a decision tree so that each layer can separate one or more classes from the rest of the classes. The process goes on and on until it determines the final category of the samples by the leaf node. For M classes of samples, we can achieve the classification of samples by training only $M - 1$ decision planes.

Different construction methods can be chosen to construct the different decision trees. If the number of two classes of nodes is equal in each classification, such a decision tree structure is called a balanced decision tree. A decision tree with an imbalanced number of nodes on both sides is called an imbalanced decision tree (see Figure 2). Obviously, in the process of classification, the imbalanced binary tree selects one class and the rest classes to be divided every time, while the balanced binary tree selects multiple classes and multiple classes to be divided each time.

For M multiclassification problems, the binary decision tree only needs to train $M - 1$ classifiers, but the OVO and directed acyclic graph need to train $M(M - 1)/2$ classifiers. In contrast, the former can save the repeated training time. At the same time, when using the binary decision tree method, the judgment of class be a single class, and there is no case that an object outputs multiple class labels or no judgment result. So, it avoids the indivisibility of the problems of the OVA and OVO algorithm. However, the binary decision tree algorithm also has its own drawbacks, and its classification effect is affected by the binary tree structure. In the binary tree classification, the “error accumulation” phenomenon often occurs; that is, the error occurring at a certain node will spread to the next layer of nodes, making the classification error of the next level further expand or even lose the meaning of classification. Therefore, the higher the node where the error occurs, that is, the earlier it occurs, the larger the scope of the error’s influence and the worse the classifier’s effect.

In general, the first step of the method of a binary tree based on SVM is to construct between-classes separability measure, then find out the two classes or two class combinations with the largest measure value, identify one or a group as a positive class and another one or a group as a negative class, and train the classification decision hyperplane through SVM; in the same way, the lower nodes are grouped and trained through the between-classes separability measure until the leaf nodes are output.

2.3. Between-Classes Separability Measure. As mentioned above, the “error accumulation” phenomenon is easy to occur in the binary tree multiclassification algorithm. In order to solve this problem better, many scholars have done a lot of work: literature [17, 18] attempt to construct a more reasonable decision tree by improving the measure of separation between classes. In reference [19], the equivalent distance is used as the separability measure. During the training, it grows from the leaves to the root. Each time the classifier is trained,

two local class clusters with the minimum equivalent distance are found for separation. In order to further improve the classification performance of decision tree nodes, literature [20] proposes an improved dual-support vector machine algorithm. Literature [21] proposes a SVM multiclassification decision tree based on the cumulative fitness genetic algorithm. This algorithm constructs a binary tree layer by layer from the root node, selects a new population, and searches for the global optimal solution to classify training samples by redefining the genetic fitness function.

For the multiclassification decision tree, the earlier the classification error occurs and the closer to the upper nodes, the worse the classification effect. Therefore, to construct a reasonable algorithm and tree structure, it is necessary to first separate the classes that are not easily misclassified. In other words, the closer the nodes are to the upper layers, the greater the difference between the two classes should be. In this way, the occurrence of the classification errors can be controlled as far as possible away from the root nodes. Therefore, separating the classes with the greatest differences as early as possible is the key to implementing the above steps. In order to achieve this goal, it is necessary to construct a reasonable between-classes separability measure.

2.3.1. The Introduction of between-Classes Separability Measure. The between-classes separability measure refers to the degree of separation between different classes. In the feature space, each type of data object corresponds to a class field, which is the minimum convex set containing such data objects. The best separation of two class fields means that there is no intersection between them. In addition, if the two classes have overlapping regions, the smaller the proportion of data elements in the overlapping region to the total data elements, the easier the two classes of data are to be separated and the better the separability is [22]. As the sample space shown in Figure 3, class 2 has the best separability because it does not intersect with the other classes. However, both class 3 and class 4 intersect with class 1 and the sample proportion of intersection area between class 4 and class 1 is larger than that between class 3 and class 1. Therefore, class 3 has a better separability than class 4. Through the above analysis of the separability between-classes, we can first separate class 2 with the best separability from the rest classes to train a SVM binary classifier; then choose class 3 with better separability and separate it from class 1 and 4; and finally, separate class 1 and 4. The imbalanced binary tree constructed in accordance with the above ideas is shown in Figure 4.

It can be seen that for the SVM multiclassification methods based on decision tree, constructing a reasonable separability measure has an important impact on the classification performance of the decision tree, and it is also the focus of current research. Therefore, it is the key to reduce the “error accumulation,” improve the classification accuracy, and improve the classification efficiency by separating the easily distinguishable classes first and reducing the error rate of the upper nodes as much as possible.

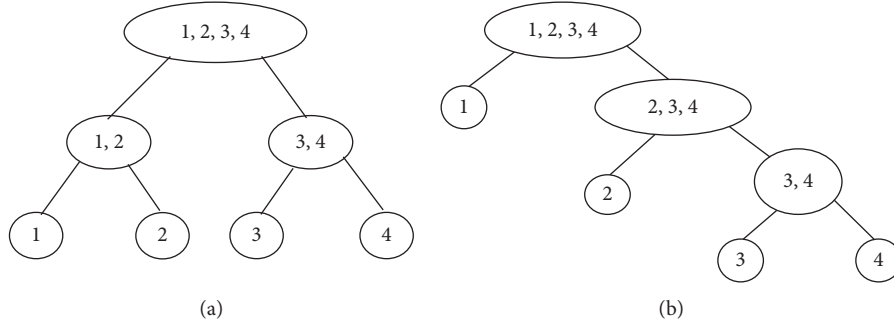


FIGURE 2: The structures of (a) a balanced binary tree and (b) an imbalanced binary tree.

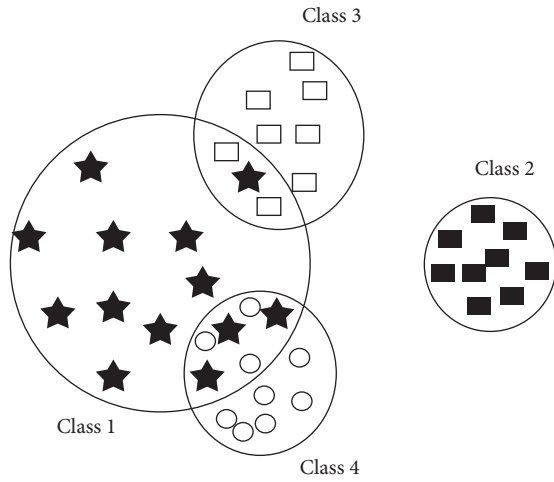


FIGURE 3: The distribution between-classes diagram.

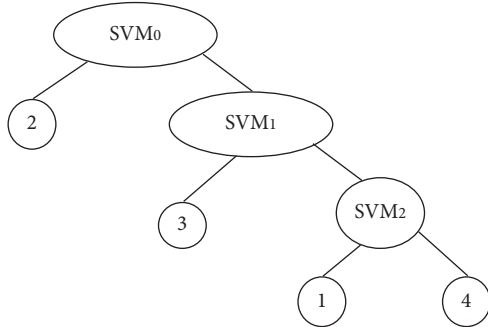


FIGURE 4: The imbalanced binary tree constructed according to the separability measure.

2.3.2. The Description of the Existing Algorithms. The binary tree classification algorithm based on SVM proposed in literature [23] uses the minimum distance method as the between-classes separability measure. This algorithm first finds the two classes of samples with the longest distance between classes to form two different datasets, while the rest of the classes classify the two sets according to the nearest principle and finally forms two categories of samples, positive and negative, for classifier training. Then, the samples of each class are recursively recurred according to the above algorithm until each output is a single category.

This algorithm is called BDT-SVM algorithm. Literature [17] defines the separation measure by “the distance between the center of two types of samples and the ratio of the variance sum of the two types of samples themselves.” The higher the value is, the higher the degree of separation is. In this paper, this algorithm is called VBDT-SVM algorithm. In literature [24], the separability measure proposed in literature [17] is used as fitness function to optimize the decision tree with the genetic algorithm. Different classification strategies are adopted according to the degree of node separability measure in classification, and better results are obtained.

The above algorithms construct the between-classes separability measure from different angles, and both of them achieve good classification effect. However, there are still some shortcomings, which are shown as follows:

- (1) The BDT-SVM algorithm only considers the distance between classes as the index of class separation degree. Although it is simple, it ignores the distribution characteristics within samples. In addition to the influence of the distance between classes, the distribution characteristics of samples also affect the degree of separation between classes. Literature [17] points out that when the distance between the class are same, the degree of separation between the classes is closely related to their own distribution, and the degree of separation between the classes is inversely proportional to the dispersion degree of the sample's own distribution. Therefore, the VBDT-SVM algorithm proposed in literature [17] defines the measure of between-classes separability as the ratio of the distance between the center point of two types of samples and the sum of the variance of the two types of samples themselves (see formula (5)). The higher the ratio is, the easier the separation of the two classes is.

$$\text{separation}_{ij} = \frac{d_{ij}}{\delta_i + \delta_j}, \quad (5)$$

where d_{ij} represents the distance between the samples of class i and class j , δ_i and δ_j represent the class variances of class i and class j , respectively (see formula (9)), and represent the distribution tightness of the two classes of samples themselves.

- (2) Formula (5) not only considers the distance of class centers also considers the class variance, which is the discrete degree of the distribution of the samples in the same class. However, the algorithm still exists deficiencies; for example, it does not consider that when the between-classes distance and class variance are same, the separability of classes also depends on the distribution tightness of one class objects relative to the other classes of objects, that is, the between-classes variance (see Definition 1).

3. The Design of IBDT-SVM Algorithm

In summary, in order to construct a more reasonable binary decision tree, it is necessary to solve the following problems:

- (1) It should construct a new between-classes separability measure and find the two classes with the greatest difference according to the between-classes separability measure at the upper node as far as possible so as to avoid bringing the errors to the nodes of the next layer.
- (2) It should construct a new building-tree scheme and a reasonable classifier as close to the root node as possible. In this way, it is necessary to carry out a second revision for each classifier in order to improve the classification accuracy and avoid the “error accumulation” to the greatest extent.

Based on the above considerations, this improved algorithm divides the samples with the greatest difference step by step according to the new between-classes separability measure to train the classifier and then classifies the rest classes according to the class-grouping-by-majority principle to form the new training samples and trains the classifier again. Thus, an improved SVM multiple classification algorithm based balanced binary decision tree, known as IBDT-SVM algorithm, is obtained.

3.1. The Improved Between-Classes Separability Measure. In order to solve the problems existing in the traditional between-classes separability measure, inspired by three

decision-making of clustering thought [25], this paper proposes between-classes separability measure about q neighbors, and the new between-classes separability measure mainly considers the following three factors as follows. (1) *The Between-Classes Variance*. Starting from the number and distance between the samples, it reflects the closeness of the relationship between a certain class of objects and its neighboring classes. It is inversely proportional to the between-classes separability measure; that is, the smaller the value, the greater the separability measure of the two classes of samples. (2) *Class Variance* δ . It reflects the compactness of distribution of samples of the class itself and is inversely proportional to the between-classes separability measure; that is, the smaller the value, the greater the separation of the class from other classes. (3) *Between-Classes Distance*. It is the distance between two class centers and is proportional to the between-classes separability measure; the greater the value is, the greater the separability of the two classes of samples is.

Definition 1 (between-classes variance). Considering one sample's q neighbors in the other neighbor class, its value indicates the degree of separation of one class objects from another class.

- (1) The degree of separation between object x_i in class C_m and Class C_k is expressed as follows:

$$\text{Sep}(x_i, C_k) = \frac{\sum_{x_j \in C_k} \alpha(x_i, x_j) \cdot \|x_i - x_j\|}{q} \quad (x_i \in C_m), \quad (6)$$

$$\text{where } \alpha(x_i, x_j) = \begin{cases} 1, & x_j \in C_k, x_j \in N_q(x_i) \\ 0, & x_j \notin C_k, x_j \in N_q(x_i) \end{cases}$$

$N_q(x_i)$ represents the nearest neighbors of the object x_i ; $q \cdot \text{Sep}(x_i, C_k)$ represents the sum of the distances between x_i and the objects that are in class C_k but also belong to $N_q(x_i)$.

- (2) Separability between all objects in class C_m and class C_k is as follows:

$$\text{sep}_{C_{mk}}(C_m, C_k) = \sum_{x_i \in C_m} \text{sep}(x_i, C_k) = \frac{\sum_{x_i \in C_m} \sum_{x_j \in C_k} \alpha(x_i, x_j) \cdot \|x_i - x_j\|}{q}. \quad (7)$$

Similarly, separability between all objects in class C_k and class C_m is as follows:

$$\text{sep}_{C_{km}}(C_k, C_m) = \sum_{x_j \in C_k} \text{sep}(x_j, C_m) = \frac{\sum_{x_j \in C_k} \sum_{x_i \in C_m} \alpha(x_j, x_i) \cdot \|x_j - x_i\|}{q}, \quad (8)$$

$$\text{where } \alpha(x_j, x_i) = \begin{cases} 1, & x_i \in C_m, x_j \in N_q(x_i) \\ 0, & x_i \notin C_m, x_j \in N_q(x_i) \end{cases}.$$

Definition 2 (class variance). Suppose class m has n_m samples and the center is μ_m , then $\mu_m = (1/n_m) \sum_{x_i \in C_m} x_i$; suppose class k has n_k samples and the center is μ_k , then $\mu_k = (1/n_k) \sum_{x_j \in C_k} x_j$; δ_m and δ_k are defined as class variances of class m and class k , respectively; that is:

$$\begin{aligned} \delta_m &= \frac{1}{n_m} \sum_{x_i \in C_m} \|x_i - \mu_m\|, \\ \delta_k &= \frac{1}{n_k} \sum_{x_j \in C_k} \|x_j - \mu_k\|. \end{aligned} \quad (9)$$

On the basis of considering the between-classes variance, class variance, and between-classes distance, it improves the separability measure function again, as defined in the following.

Definition 3 (between-classes separability measure). The degree of separation between class C_m and class C_k is defined as follows:

$$\begin{aligned} \text{separation}(C_m, C_k) &= \max \left\{ \frac{d_{mk}}{\delta_m + \text{sep}C_{mk}}, \frac{d_{mk}}{\delta_k + \text{sep}C_{km}} \right\} \\ &= \max \left\{ \frac{d_{mk}}{\delta_m + \sum_{x_i \in C_m} \text{sep}(x_i, C_k)}, \frac{d_{mk}}{\delta_k + \sum_{x_j \in C_k} \text{sep}(x_j, C_m)} \right\}, \end{aligned} \quad (10)$$

where $d_{mk} = \|\mu_m - \mu_k\|$ represents the distance between the centers of class C_m and class C_k , that is, the between-classes distance.

3.2. Class Grouping Algorithm Based on the Principle of Class-Grouping-by-Majority. The IBDT-SVM multiclassification algorithm proposed in this paper improves the between-classes separability measure on the basis of considering the between-classes distance, class variance, and between-classes variance. According to the improved between-classes separability measure, the two classes with the highest separability measure are first found and the classification model is trained. Then, it uses the class-grouping-by-majority principle [26] to group the other classes into these two groups and use them as the training samples to retrain the classifier. This algorithm loops on each decision surface until each output is a leaf node, that is, a single sample point. This method can ensure that all classes can be separated as far as possible at each classification node and the classification of the rest classes as reasonable as possible. For the data with uneven or sparse distribution, the error caused by the classification method of minimum distance between class centers can be avoided.

3.2.1. The Introduction of the Class-Grouping-by-Majority Principle. Now take the classification of two-dimensional samples of class “1,” “2,” and “3” (see Figure 5) as an example and introduce the principle of class-grouping-by-majority. If

class “1” and class “2” are used as training sets to train the SVM classifier, the samples after classification are scattered on both sides of the initial classifier (see Figure 5(a)). For samples of class “3,” they are distributed to both sides of the initial classifier, but the class “2” side contains most of the samples of class “3.” According to the class-grouping-by-majority algorithm, all examples of class “3” and class “2” were reclassified into one class, and the examples of class “1” are classified as another class samples. In this way, the SVM classifier is retrained and the final classifier is obtained (see Figure 5(b)). The structure of the final binary decision tree is shown in Figure 5(c).

3.2.2. Class Grouping Algorithm Based on the Principle of Class-Grouping-by-Majority. The idea of the IBDT-SVM algorithm is that on each layer of the decision tree first finding the two classes based on the improved maximum between-classes separability measure mentioned above which have the greatest difference, namely, the two classes that are easiest to be separated. It denotes the two classes as class \max_1 and class \max_2 . It takes the two classes’ samples as the positive and negative samples to train the classifier, denoted as old-classifier. Then, based on the principle of class-grouping-by-majority, it groups the other classes closer to the two classes to form the two new classes: new- \max_1 and new- \max_2 . Then, it takes the two classes: class new- \max_1 and class new- \max_2 as the positive and negative samples to train the SVM classifier again, denoted as new-classifier. The cycle continues until all classification results are single class.

Take a 4-classes classification problem as an example (see Figure 6). Suppose that among the four classes, class 1 and class 2 have the maximum separability measure, we find out class 1 and class 2 first and take them as the positive and negative samples, respectively, to train the classifier, that is, old-classifier. Then, it puts the training samples of class 3 and class 4 into the classifier for testing according to the class-grouping-by-majority principle. There are four possibilities:

- (1) If most of the samples of class 3 are on the positive side of the old-classifier, while most of the samples of class 4 are on the negative side of the old-classifier, then the classified result is {1, 3}, {2, 4} (see Figure 6(a))
- (2) If most of the samples of class 4 are on the positive side of the old-classifier, while most of the samples of class 3 are on the negative side of old-classifier, the classification result is {1, 4}, {2, 3} (see Figure 6(b))
- (3) If most of the samples of the two classes such as class 3 and class 4 are all biased to the positive side of the old-classifier, then the classification result is {1, 3, 4}, {2} (see Figure 6(c))
- (4) If most of the samples of the two classes such as class 3 and class 4 are all biased to the negative side of the old-classifier, then the classification result is {1}, {2, 3, 4} (see Figure 6(d))

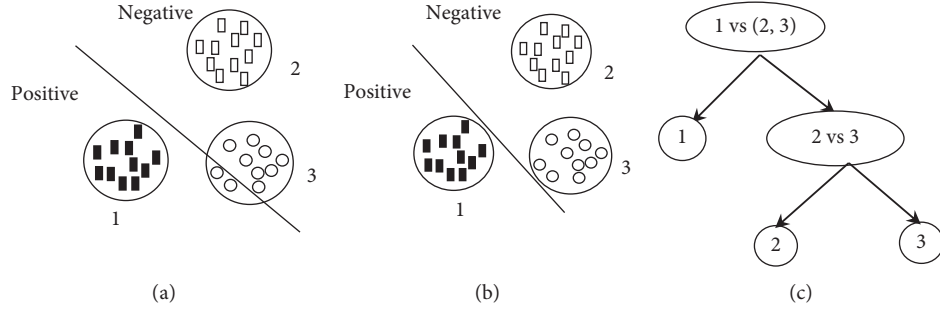


FIGURE 5: An example of the class-grouping-by-majority algorithm.

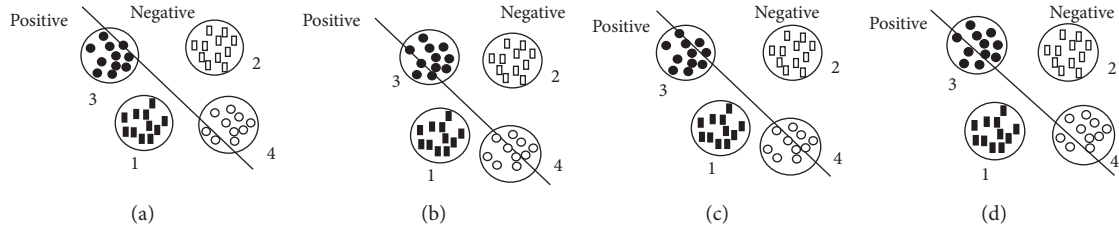


FIGURE 6: The old-classifier obtained from the initial training.

For the above situations, the reclassified samples need to be treated as the positive and negative objects to retrain the classifier, and the new-classifier is obtained (see Figure 7).

3.3. IBDT Algorithm. The idea of the improved binary decision tree algorithm called as IBDT algorithm in this paper is as follows.

Assume that there are M classes of input data, and class i contains n_i samples, then $n_1 + n_2 + \dots + n_i = n$. The IBDT algorithm process is as follows.

Step 1. Set the initial q value and calculate the separability measure values of every two classes of samples in M class which is calculated according to formulas (6)–(10), namely, the value of separation.

Step 2. Set the two classes which have the maximum between-classes separability measure be class max 1 and class max 2 and then find them out;

Step 3. The classifier is trained by using class max 1 and class max 2 as the training samples, denoted as old-classifier.

Step 4. According to the principle of the class-grouping-by-majority, the remaining class samples are classified into class max 1 and class max 2, respectively, and form two major classes $\{\text{max}_1, M_{i_1}, M_{i_2}, \dots, M_{i_j}\}$ and $\{\text{max}_2, M_{i_{j+1}}, M_{i_{j+2}}, \dots, M_{i_{M-2}}\}$, and then the two major class samples are marked as the positive samples and negative samples, training the SVM classifier to form the classifier, denoted as new-classifier.

Step 5. Repeat the above operations until the each outputting sample is labeled as a single category.

4. The Numerical Experiments and Results

We used five multiclass datasets in the UCI database such as Segmentation, Statlog, Iris, Breast tissue, and Page Blocks to verify the classification accuracy of the improved algorithm proposed in this paper-IBDT-SVM. We use the tenfold cross-validation algorithm to select the training set and the test set. Each time a subset is selected as the test set, and the average cross-validation recognition accuracy of 10 times is taken as the last result.

The above five datasets are classified using the following algorithms including OVO, OVA, BDT-SVM, VBDT-SVM algorithm, and the IBDT-SVM algorithm proposed in this paper, and the classification accuracy is compared.

In this study, we used five multiclass datasets from the UCI database, with different sample sizes, number of attribute values, and number of categories. The full name and corresponding abbreviation of the dataset used in this experiment are shown in Table 1, and the characteristics of the above multiclass datasets are shown in Table 2, including data set name, sample number, number of attributes of samples (data dimension), and number of categories.

The classification accuracy of 10 experiments using the OVO algorithm, OVA algorithm, BDT-SVM algorithm, VBDT-SVM algorithm, and IBDT-SVM algorithm for the above 5 datasets is shown in Tables 3–7). The average classification accuracy of the above results is shown in Table 8). It can be seen that the IBDT-SVM algorithm has a good classification effect.

The comparison results of the classification accuracy of OVO algorithm, OVA algorithm, BDT-SVM algorithm, VBDT-SVM algorithm, and IBDT-SVM algorithm for the above five different datasets in 10 experiments are shown in Figures 8(a)–8(e). It can be seen from Figure 8, compared with the other four multiclassification algorithms, the IBDT-

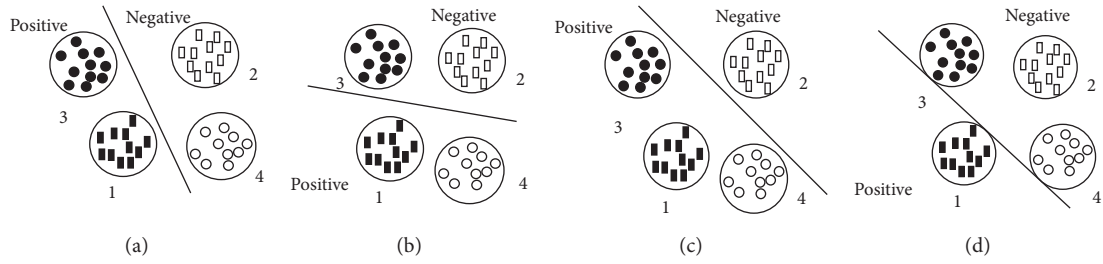


FIGURE 7: The retrained new-classifier after classification according to the class-grouping-by-majority principle.

TABLE 1: The names and abbreviations of the datasets.

Dataset ID	Dataset names	Dataset abbreviations
1	Image segmentation	Segmentation
2	Statlog (Landsat satellite)	Statlog
3	Iris	Iris
4	Breast tissue	Breast tissue
5	Page blocks classification	Page Blocks

TABLE 2: The information of the datasets in this study.

Dataset ID	Dataset abbreviations	Sample sizes	Number of attribute values	Number of categories
1	Segmentation	2310	19	7
2	Statlog	6435	36	7
3	Iris	150	4	3
4	Breast tissue	106	10	6
5	Page Blocks	5473	10	5

TABLE 3: The classification accuracy of each multiclassification algorithm for the dataset “Segmentation” in the 10 experiments.

	1	2	3	4	5	6	7	8	9	10
OVO	0.923	0.961	0.961	0.957	0.957	0.965	0.961	0.957	0.957	0.960
OVA	0.880	0.940	0.944	0.935	0.935	0.948	0.939	0.935	0.939	0.943
BDT-SVM	0.944	0.987	0.972	0.987	0.974	0.985	0.981	0.987	0.978	0.991
VBDT-SVM	0.974	0.976	0.989	0.981	0.983	0.981	0.987	0.987	0.981	0.985
IBDT-SVM	0.987	0.989	0.991	0.991	0.991	0.972	0.978	0.978	0.985	0.985

TABLE 4: The classification accuracy of each multiclassification algorithm fort the dataset “Statlog” in the 10 experiments.

	1	2	3	4	5	6	7	8	9	10
OVO	0.967	0.948	0.906	0.833	0.945	0.943	0.969	0.962	0.945	0.896
OVA	0.922	0.891	0.857	0.757	0.878	0.888	0.923	0.916	0.887	0.839
BDT-SVM	0.961	0.961	0.955	0.848	0.953	0.956	0.947	0.946	0.941	0.935
VBDT-SVM	0.961	0.939	0.956	0.950	0.957	0.952	0.952	0.950	0.944	0.945
IBDT-SVM	0.940	0.946	0.957	0.957	0.952	0.961	0.953	0.945	0.955	0.952

TABLE 5: The classification accuracy of each multiclassification algorithm for the dataset “Iris” in the 10 experiments.

	1	2	3	4	5	6	7	8	9	10
OVO	1.00	0.933	0.970	1.00	0.933	0.933	0.933	1.00	1.00	1.00
OVA	0.933	0.867	1.000	0.933	0.933	0.933	0.867	0.933	1.0	1.0
BDT-SVM	0.933	0.933	0.967	0.90	0.967	0.933	0.967	0.900	0.900	0.967
VBDT-SVM	0.967	0.967	1.000	0.933	1.000	0.933	0.900	0.933	0.900	0.900
IBDT-SVM	0.972	1.000	0.967	0.980	0.967	0.967	0.967	0.933	0.967	0.980

TABLE 6: The classification accuracy of each multiclassification algorithm for the dataset “Breast tissue” in the 10 experiments.

	1	2	3	4	5	6	7	8	9	10
OVO	0.286	0.385	0.417	0.333	0.364	0.300	0.556	0.556	0.625	0.625
OVA	0.286	0.308	0.333	0.333	0.273	0.300	0.222	0.333	0.375	0.375
BDT-SVM	0.810	0.762	0.762	0.667	0.571	0.619	0.714	0.667	0.810	0.714
VBDT-SVM	0.571	0.762	0.571	0.714	0.571	0.571	0.619	0.429	0.571	0.762
IBDT-SVM	0.762	0.714	0.714	0.714	0.810	0.810	0.619	0.667	0.857	0.571

TABLE 7: The classification accuracy of each multiclassification algorithm for the dataset “Page Blocks” in the 10 experiments.

	1	2	3	4	5	6	7	8	9	10
OVO	0.944	0.947	0.936	0.932	0.942	0.949	0.929	0.885	0.936	0.938
OVA	0.942	0.954	0.940	0.936	0.942	0.948	0.934	0.879	0.934	0.945
BDT-SVM	0.915	0.900	0.911	0.915	0.906	0.926	0.923	0.915	0.900	0.909
VBDT-SVM	0.910	0.905	0.914	0.888	0.898	0.917	0.916	0.919	0.921	0.899
IBDT-SVM	0.928	0.918	0.928	0.950	0.930	0.928	0.950	0.930	0.928	0.960

TABLE 8: The comparison of average classification accuracy of the five multiclassification algorithms.

Data sets	OVO	OVA	BDT-SVM	VBDT-SVM	IBDT-SVM
Segmentation	0.956	0.934	0.979	0.982	0.985
Statlog	0.931	0.876	0.940	0.950	0.952
Iris	0.970	0.940	0.937	0.943	0.970
Breast tissue	0.445	0.314	0.710	0.614	0.724
Page Blocks	0.934	0.935	0.912	0.909	0.935

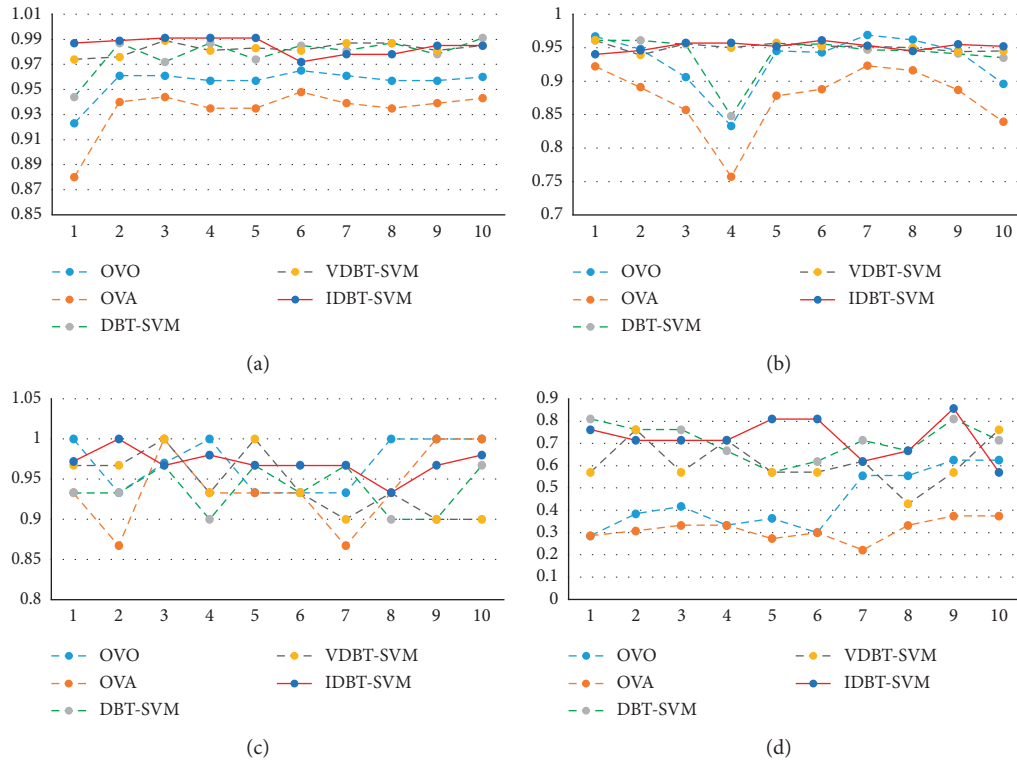


FIGURE 8: Continued.

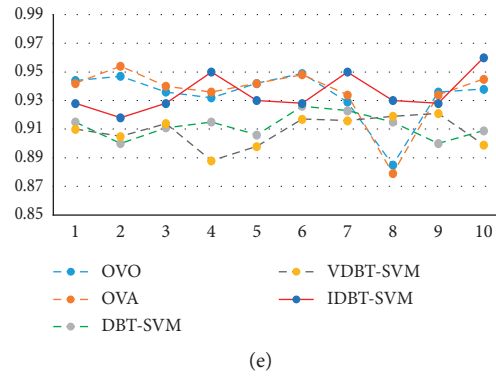


FIGURE 8: The classification accuracy of 5 multiclassification algorithms for dataset (a) Segmentation, (b) Statlog, (c) Iris, (d) Breast tissue, and (e) Page Blocks.

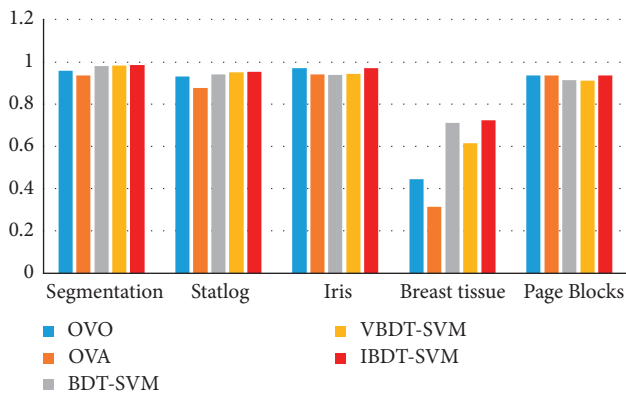


FIGURE 9: The average classification precision of 5 multiclassification algorithms for each dataset.

SVM algorithm proposed in this paper has better stability and higher classification accuracy. It shows the average classification accuracy of the above five algorithms by a bar chart (see Figure 9).

5. Conclusion

In this paper, an improved multiclassification algorithm based on the balanced binary decision tree, called IDBT-SVM algorithm, is proposed. This algorithm improves the original BDT-SVM algorithm's between-classes separability measure so that the calculation of separability between-classes not only considers the distance between classes but also considers the class variance and the variance between-classes. According to the improved between-classes separability measure, the two classes with the greatest difference will be used to train the old classifier. In addition, considering the classification error of nonuniform data or sparse data using the minimum center distance classification method, the class-grouping-by-majority principle is used to classify the rest samples, and the new-classifier is trained again so as to ensure the better classification effect.

The algorithm proposed by this paper is simple and easy, its need to construct the classifier of the number is less, and the experiments show that it can improve the classification accuracy than the original binary tree classification algorithm. In the future work, we also can consider to construct a

better algorithm or decision tree structure, such as considering through the improved SVM algorithm to further optimize the decision tree classification mechanism; constructing the SVM decision tree classification based on the distance between each center and root node; or combining the advantages of other algorithms to solve the classification accuracy of multiclassification, such as forming a multiclassification model through rough set theory with complementary advantages.

Data Availability

The imbalanced dataset data used to support the findings of this study are deposited in UCI dataset and KEEL dataset, and they are available openly (the URL of the UCI dataset is <http://archive.ics.uci.edu/ml/datasets.php> and the URL of the KEEL dataset is <https://sci2s.ugr.es/keel/dataset.php>).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (project no. 61976244) and the Special project of Shaanxi Provincial Department of Education (project no. JK190646).

References

- [1] W. H. Xie, G. Q. Liang, Z. H. Dong et al., "An improved oversampling algorithm based on the samples' selection strategy for classifying imbalanced data," *Mathematical Problems In Engineering*, vol. 2019, Article ID 3526539, 13 pages, 2019.
- [2] P. X. Yang and P. C. Zhang, "Distributed generation system interconnection protection based on multi-class SVM," *Automation of Electric Power Systems*, vol. 39, no. 22, pp. 110–151, 2015.
- [3] D. Li, W. Yang, and S. Wang, "Classification of foreign fibers in cotton lint using machine vision and multi-class support vector machine," *Computers and Electronics in Agriculture*, vol. 74, no. 2, pp. 274–279, 2010.

- [4] E. J. Bredensteiner and K. P. Bennett, "Multicategory classification by support vector machines," *Computational Optimization*, vol. 12, no. 3, pp. 53–79, 1999.
- [5] E. Allwein, R. E. Schapire, and Y. Singer, "Reducing multiclass to binary: a unifying approach for margin classifiers," *Journal of Machine Learning Research*, vol. 1, no. 2, pp. 113–114, 2000.
- [6] R. Debnath, N. Takahide, and H. Takahashi, "A decision based one-against-one method for multi-class support vector machine," *Pattern Analysis and Application*, vol. 7, no. 2, pp. 164–175, 2004.
- [7] R. Rifkin and A. Clautau, "In defense of one-vs-all classification," *Journal of Machine Learning Research*, vol. 2004, no. 5, pp. 101–141, 2004.
- [8] L. Xu, G. Z. Zhao, and H. Gu, "Novel one-vs-rest classifier based on SVM and multi-spheres," *Journal of Zhejiang University*, vol. 43, no. 2, pp. 303–308, 2009.
- [9] J. C. Platt, N. Cristianini, and J. Shawe-Taylor, "Large margin DAGs for multi classification," *Advances in Neural Information Processing Systems*, vol. 12, pp. 547–553, 2000.
- [10] C. Zhang, Z. F. Ma, X. X. Niu et al., "A method based on dynamic sequence analysis of an API and DAG-SVM multiclass support vector machines of unknown virus detection," *Journal of Chinese Computer Systems*, vol. 33, no. 12, pp. 2724–2728, 2012.
- [11] O. Pujol, P. Radeva, and J. Vitrià, "Discriminant ECOC: a heuristic method for application dependent design of error correcting output codes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 6, pp. 1007–1012, 2006.
- [12] T. G. Dietterich and G. Bakiri, "Solving multiclass learning problems via error-correcting output codes," *Journal of Artificial Intelligence Research*, vol. 2, pp. 263–286, 1995.
- [13] D. Gjorgjevikj, "A multi-class SVM classifier utilizing binary decision tree," *Informatica*, vol. 33, no. 2, pp. 225–233, 2009.
- [14] Z. H. Diao, C. J. Zhao, X. Y. Guo et al., "A new SVM multi-class classification algorithm based on balance decision tree," *Control and Decision*, vol. 26, no. 1, pp. 149–156, 2011.
- [15] R. Y. Chen and J. F. Jiang, "Pose-varied face recognition based on neural network and hierarchical support vector machines," *Computer Engineering*, vol. 32, no. 24, pp. 210–215, 2006.
- [16] B. Lu and J. B. Yang, "A strategy of multi-level classification based on SVM," *Computer Engineering*, vol. 31, no. 1, p. 113, 2005.
- [17] S. Y. Xia, H. Pan, and L. Z. Jin, "Multi-class SVM method based on non-balanced binary tree," *Computer Engineering and Applications*, vol. 45, no. 17, pp. 167–169, 2009.
- [18] X. R. Li, G. Y. Zhao, and L. Y. Zhao, "Design of decision-tree-based support vector machines multi-class classifier based on vector project," *Journal of Control and Decision*, vol. 23, no. 7, pp. 745–750, 2008.
- [19] R. Wang and J. Chen, "Application of support vector machine-based decision tree classifier in rotor fault diagnosis," *Journal of Vibration and Impact*, vol. 29, no. s, pp. 258–260, 2010.
- [20] J. Y. Xie, B. Q. Zhang, and W. Z. Wang, "A partial binary tree algorithm for multiclass classification based on twin support vector machines," *Journal of the Nanjing University (Natural Science)*, vol. 47, no. 4, pp. 354–363, 2011.
- [21] Q. S. Zhu and K. Cheng, "SVM decision-tree multi-classification strategy based on genetic algorithm with cumulative fitness," *Computer Application Research*, vol. 33, no. 291, pp. 64–67, 2016.
- [22] F. Y. Sun, *Study on Multi-Class Classification Method of Support Machine Based on Decision Tree*, pp. 10–18, Northeast Normal University, Changchun, China, 2015.
- [23] E. Y. Wu and J. Lv, "Research on multiclass classification algorithm based on binary tree SVM," *Journal of Chongqing Normal University (Natural Science)*, vol. 33, no. 3, pp. 102–106, 2016.
- [24] S. F. Ding, J. G. Sun, and D. L. Chen, "Improved SVM decision-tree and its application in remote sensing classification," *Application Research of Computers*, vol. 29, no. 3, pp. 1146–1148, 2012.
- [25] H. Yu and C. K. Mao, "Automatic three-way decision clustering algorithm based on k-means," *Computer Applications*, vol. 36, no. 8, pp. 2061–2065, 2016.
- [26] P. Kantavat, B. Kijsirikul, P. Songsiri, K.-I. Fukui, and M. Numao, "Efficient decision trees for multi-class support vector machines using entropy and generalization error estimation," *International Journal of Applied Mathematics and Computer Science*, vol. 28, no. 4, pp. 705–717, 2018.