



A New Comprehensive Evaluation Method for Water Quality: Improved Fuzzy Support Vector Machine

Wei Shan ^{1,2,*}, Shensheng Cai¹ and Chen Liu³

- ¹ School of Economics and Management, Beihang University, Beijing 100191, China; caishensheng@126.com
- ² Key Laboratory of Complex System Analysis and Management Decision, Ministry of Education, Beijing 100191, China
- ³ Business School, University of Shanghai for Science and Technology, Shanghai 200093, China; liuchen.chn@hotmail.com
- * Correspondence: shanwei@buaa.edu.cn

Received: 3 September 2018; Accepted: 19 September 2018; Published: 21 September 2018



Abstract: With the pressure of population growth and environmental pollution, it is particularly important to develop and utilize water resources more rationally, safely, and efficiently. Due to safety concerns, the government today adopts a pessimistic method, single factor assessment, for the evaluation of domestic water quality. At the same time, however, it is impossible to grasp the timely comprehensive pollution status of each area, so effective measures cannot be taken in time to reverse or at least alleviate its deterioration. Thus, the main propose of this paper is to establish a comprehensive evaluation model of water quality, which can provide the managers with timely information of water pollution in various regions. After considering various evaluation methods, this paper finally decided to use the fuzzy support vector machine method (FSVM) to establish the model that is mentioned above. The FSVM method is formed by applying the membership function to the support vector machine. However, the existing membership functions have some shortcomings, so after some improvements in these functions, a new membership function is finally formed. The model is then tested on the artificial data, UCI dataset, and water quality evaluation historical data. The results show that the improvement is meaningful, the improved fuzzy support vector machine has good performance, and it can deal with noise and outliers well. Thus, the model can completely solve the problem of comprehensive evaluation of water quality.

Keywords: water quality; fuzzy support vector machine; membership function; comprehensive evaluation

1. Introduction

Water is an essential resource for human survival and development, and it has become more and more important because of the growth of population and the deterioration of environmental pollution [1–5]. Therefore, a model that can distinguish water quality is critical. It helps managers to rationally allocate and utilize water resources, and brings a more comprehensive understanding of water pollution in various areas at the same time.

Existing water quality assessment methods can be broadly divided into three categories: traditional assessment method, evaluation method based on fuzzy mathematics and machine learning method. For the first one, traditional water quality assessment methods such as the single factor assessment, grading score method, function evaluation method, etc. used a series of calculations to obtain a comprehensive score to evaluate water quality [6]. However, due to the non-determinism and non-linearity of water pollution, traditional methods cannot accurately describe this complex pollution process. For the second one, since the classification boundaries and pollution levels are both fuzzy



phenomena, fuzzy mathematics was applied to water quality assessment. Among them, the fuzzy comprehensive evaluation [7–9] and the gray clustering method [10,11] were widely used. Besides, Yun and Zou combined these two methods for water quality evaluation [12]. However, on the one hand, it takes time and effort for experts to give scores, and on the other hand, the choice of whitening function varies from person to person, so these methods that are based on fuzzy mathematics are difficult to be used widely. For the third one, because today's machine learning methods can solve nonlinear problems well, these algorithms have been widely used in the field of water quality evaluation. In terms of neural networks, it has been widely used in the evaluation of water quality [13–16]. However, the neural network algorithm requires a significant amount of samples for training. In contrast, support vector machine (SVM) has good generalization ability and unique advantage in solving the classification problem of nonlinear high-dimensional modes in the case of a small number of samples [17].

Vladimir Vapnik originally proposed the SVM for the two-class classification problem [18]. It has been applied to classification and prediction problems in various fields such as medicine, engineering, and education. However, as the SVM is applied to many fields, the performance of standard SVM method is gradually insufficient to meet our needs. Therefore, many researchers have optimized it to further improve the performance of this evaluation model. In terms of parameter optimization, ant colony algorithm [19], genetic algorithm [20], and particle swarm optimization algorithm [21], etc. were used to help the SVM search for the optimal parameters g and c faster. In addition, in terms of model building, Fei and Liu provided a new binary tree-based SVM algorithm, which can improve classification efficiency [22]. Liu et al. proposed an efficient self-adaption instance selection algorithm to reconstruct the training set of support vector machine [23]. Li et al. used Adaboost to improve SVM [24]. Suykens et al. proposed a least squares support vector machine, which can reduce the computational complexity [25].

Therefore, SVM is widely used in water quality evaluation, and its performance is constantly being improved. Zhou et al. used SVM to evaluate the water quality data of Wei River, and proposed a self-adaptive parameter optimization method using float genetic algorithm [26]. Chen et al. used SVM to evaluate the groundwater quality of Yangmaowan Irrigated Area [27]. Dai also used intelligent genetic algorithm to select the parameters of the least square support vector machine (LS-SVM), and then used this model to classify and predict the water quality of the Changjiang River [28].

However, this paper finds that the problem of noise points that may be formed during water quality evaluation has not been solved. Fortunately, the fuzzy support vector machine method (FSVM) was proposed. It can solve the problem of noise and outliers by applying the concept of membership function in support vector machine [29]. Applying it to water quality assessment can solve this problem.

Besides, the core of FSVM, the membership function, is also constantly being improved to better identify noise. Ren constructed a new membership function through the geometric mean of two membership functions, and found that this method can improve the classification performance [30]. Wu defined an adjustment factor by classifying hyperplanes and classification intervals [31]. Xu proposed two definition methods of membership function, one was to use two intra-surface hyperplanes to define membership degree, and the other was to consider both the distance-based membership function and the new compactness-based membership function [32]. However, after analyzing several membership functions in the above existing articles, this paper finds that these functions have some problems in some cases.

Therefore, the main purposes of this study are (1) to form a new membership function by improving the above defects; (2) to construct the FSVM water quality evaluation model that is based on this membership function; and, (3) to verify the performance of the above model and apply it to the real cases. This study aims to establish a reasonable and efficient comprehensive model of water quality assessment, by which managers can understand the timely water pollution information of each region.

2. FSVM Methodology

2.1. Data Preprocessing

2.1.1. Data Imbalance

For the water quality assessment, in order to strengthen the management of surface water environment, prevent water pollution, and protect human health, the "Surface Water Environmental Quality Standard" (GB3838-2002) promulgated by the State Environmental Protection Administration of China is used as the national quality standard for water quality assessment in China. According to it, the water quality is divided into five categories from I to V. But, in reality, some indicator values of a small amount of areas exceed the value of this evaluation table, which is evaluated as inferior V. Obviously, the data of the inferior V class is less than the other classes. As a result, the dataset is not balanced.

Therefore, this problem needs to be solved by oversampling or undersampling method. In 2002, Chawla et al. proposed the synthetic minority over-sampling technique method (SMOTE), which solves the data imbalance problem by randomly selecting a point from the k neighbors of one sample in the minority class and generating a new sample between the original sample and this selected sample [33]. However, this method may also cause additional noise, so the modified approach (MSMOTE) was proposed by Hu et al. in 2009 [34]. It first divides the minority class samples into safety samples, boundary samples and potential noise, and then oversamples the safety samples. In general, MSMOTE can be used to solve the data imbalance problem well, but sometimes there are several minority classes at the same time, then the problem can only be solved by the undersampling method.

Of course, after data balancing, the accuracy may not be improved, sometimes it would even be slightly reduced on the contrary. For the training data of water quality assessment, it is very likely that the amount of data in the inferior V is very small. However, there are many data belonging to class II or class III. After all, the excellent water resources will only be a minority, as well as the severely polluted area. Then, if the classifier for the class II and inferior class V judges that all the samples belong to the former, the accuracy rate can be very high, but, in fact, the popularity of this model is very poor, only in its own testing set has a high accuracy. Therefore, once the above methods are used to balance the dataset, it is likely to lead to a decrease in classification accuracy, but, in fact, the promotion degree of the model will be increased.

2.1.2. Data Normalization

The water quality data has different dimensions for different indicators. For example, the pH value generally belongs to the range of 6–9, while the ammonia nitrogen value is mostly between 0 and 1. So, the data needs to be normalized first. In this paper, the mapminmax function in Matlab is used for normalization, and all data is normalized to (0,1). The formula that is used by mapminmax is shown in Equation (1).

$$x_{i}' = \frac{(\max x_{i}' - \min x_{i}')(x - \min x_{i})}{\max x_{i} - \min x_{i}} + \min x_{i}'$$
(1)

where the value of max x_i' and min x_i' need to be set. Here as the range of normalization is (0, 1), we have max $x_i' = 1$, min $x_i' = 0$.

2.2. Basic Model Selection

According to the "Surface Water Environmental Quality Standards" (GB3838-2002), the classification criteria of three indicators are shown in Table 1. Now, suppose that the situation is simplified into a two-dimensional plane, then the location of samples belonging to class I and II is shown in Figure 1. The points in lower left and upper right areas are unquestionable standard water quality, but in reality, there must be sample points in the lower right and upper left areas of the figure. For these points, the value of one indicator satisfies the Class I standard, and the value of the other indicator is not satisfied.

This situation cannot be evaluated only by the water quality standard. Therefore, it is necessary for the experts to evaluate them according to the actual situation. Because of the opinion differences, noise that may interfere with the classification model will be generated. Therefore, the water quality evaluation model of this paper will be constituted by the FSVM method to reduce noise.

Category	DO (≥)	$CODmn (\leq)$	NH3-N (≤)
Ι	7.5	2	0.15
II	6	4	0.5
III	5	6	1
IV	3	10	1.5
V	2	15	2
Inferior V	<2.0	>15	>2.00
		· · · · · · · · · · · · · · · · · · ·	
0 0.05	0.1 0.15 0.2	0.25 0.3 0.35 NH ₃ -N	0.4 0.45 0.5

Table 1. Part of the surface water quality standard.

Figure 1. Part of the hypothetical water quality sample points of class I and class II.

Why can FSVM remove the effects of noise points? In the standard SVM model, the classification problem may be overfitting at the beginning, that is, the model requires a perfect separation of the two types of samples, which would result in the situation of Figure 2b instead of Figure 2a. Therefore, it is necessary to relax the original condition of $y_i(w^T z_i + b) \ge 1$ by the slack variable, that is, the value of $y_i(w^T z_i + b)$ of some points can be allowed to be less than 1, but this relaxation still requires a certain limit. Therefore, the sum of all the slack variables ε_i needs to be kept to a minimum. At the same time, the penalty parameter *C* is applied to control this degree of rigor. So, the Equation (2) is formed.

$$\min_{\substack{w,b,\varepsilon}\\w,b,\varepsilon} \left(\frac{1}{2}w^T w + C\sum_{i=1}^n \varepsilon_i\right) \\
\text{s.t.} \quad y_i(w^T z_i + b) \ge 1 - \varepsilon_i \\
\varepsilon_i \ge 0 \qquad i = 1, \dots, n$$
(2)



Figure 2. Example of overfitting. (a) Normal situation; (b) Situation of overfitting.

But, since *C* is a constant, the SVM model gives the same degree of punishment to each sample after softening the boundary. Therefore, the FSVM specifically gives different importance to different samples to solve this problem, that is, transform the sample set U by the membership degree S_i to $U' = \{(x_i, y_i, S_i), i = 1, 2, ..., n\}$, where $x_i \in R^m$, $y_i \in \{+1, -1\}$, $S_i \in (0, 1)$. Thus, the previous SVM model is changed into Equation (3).

$$\min_{\substack{w,b,\varepsilon} \\ w,b,\varepsilon} \left(\frac{1}{2} w^T w + C \sum_{i=1}^n S_i \cdot \varepsilon_i \right)$$

s.t. $y_i \left(w^T z_i + b \right) \ge 1 - \varepsilon_i$
 $\varepsilon_i \ge 0 \qquad i = 1, \dots, n$ (3)

Convert the above problem into Equation (4) while using the Lagrangian function:

$$\min_{w,b,\varepsilon} \max_{a_i \ge 0} \sum_{\beta_i \ge 0} L(w,b,\alpha,\beta) \quad i = 1,\dots,n$$

$$L(w,b,\alpha) = \frac{1}{2}w^T w + C \sum_{i=1}^n S_i \cdot \varepsilon_i + \sum_{i=1}^n \alpha_i \left[1 - \varepsilon_i - y_i \left(w^T z_i + b \right) \right] + \sum_{i=1}^n \beta_i \cdot (-\varepsilon_i)$$
(4)

Then, convert it to dual problem:

$$\max_{a_i \ge 0} \min_{w,b,\varepsilon} \left\{ \frac{1}{2} w^T w + C \sum_{i=1}^n S_i \cdot \varepsilon_i + \sum_{i=1}^n \alpha_i \left[1 - \varepsilon_i - y_i \left(w^T z_i + b \right) \right] + \sum_{i=1}^n \beta_i \cdot (-\varepsilon_i) \right\} \quad i = 1, \dots, n$$
(5)

Since $\frac{\partial L}{\partial \varepsilon} = S_i \cdot C - \alpha_i - \beta_i = 0$, the original form becomes:

$$\max_{0 \le a_i \le C} \min_{\beta_i = S_i \cdot C - \alpha_i} \min_{w, b, \varepsilon} \{ \frac{1}{2} w^T w + \sum_{i=1}^n \alpha_i \Big[1 - y_i \Big(w^T z_i + b \Big) \Big] \} \quad i = 1, \dots, n$$
(6)

Since $\frac{\partial L}{\partial b} = 0$ and $\frac{\partial L}{\partial w_i} = 0$, the above formula can be reduced to a problem only containing the unknown number α_i :

$$\min_{a_i} \left(\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j z_i^T z_j - \sum_{i=1}^n \alpha_i \right) \\
s.t. \sum_{i=1}^n \alpha_i y_i = 0 \\
0 \le a_i \le S_i \cdot C \quad i = 1, \dots, n \\
w = \sum_{i=1}^n \alpha_i y_i z_i, \ \beta_i = S_i \cdot C - \alpha_i \quad i = 1, \dots, n$$
(7)

It can be seen that since the *C* value is the same for all the sample points, the points whose S_i is larger are less likely to be misclassified, and the points whose S_i is smaller have less effect on the formation of the optimal hyperplane. Therefore, by assigning different S_i values to different sample points, the influence of sample noise can be reduced.

2.3. Parameter Optimization

The penalty parameter C indicates the degree of punishment for the misclassified samples, so the larger the C is, the smaller the error would be. However, if C is too large, this will lead to overfitting, so a suitable parameter *C* is needed. In this paper, the RBF kernel function is selected, so the size of the parameter *g* will affect the complexity of the optimal classification plane. In terms of parameter optimization, this paper adopts the grid optimization method. Each time that a set of (C, g) values are determined to be run in the model, and finally, all of the areas in the grid are traversed to obtain the best *C* and *g*.

2.4. Cross-Validation

For a classification model, after training with historical data, samples of unknown results can be put into the model to obtain prediction results. But, there is no way to verify the performance of this classifier itself. Thus, this sample set of known results is generally divided into two parts, one part is still used to train the model and the other part is used to test the performance of the trained model [35].

The commonly used cross-validation methods are as follows:

- (1)K-cv: The data is divided into k groups, then each time one of them is selected for testing, and the remaining (k-1) groups are used to train the model. This process will be repeated k times. Finally, the evaluation result is generated by taking the average of all the results.
- (2) Loo-cv: Suppose there are *n* samples, each time (n - 1) samples are used for training, and the remaining one is used as a test set. The above process is repeated *n* times, and the final result is the average of all values. Since almost every sample is used for training each time, this method leaves almost no information and the results are more reliable. However, because it is repeated too many times, the method is time consuming.

2.5. Multi-Classification Model

Since the water quality data is divided into six categories, the structure of model used in this paper is as shown in Figure 3. First, step (1) is performed, that is, each of the two types of data in the training set is separately put into the classifiers 1–15 for training. Then, step (2) is performed, that is, a new sample of an unknown result is separately put into 15 classifiers for evaluation. Finally, step (3) is to summarize the classification result of all classifiers to obtain the final prediction result.

Step (1): The two-class classification model uses the FSVM method, and the training dataset is put into the model to calculate the values of *w* and *b*.

Step (2): For the sample x with the unknown result, since $w = \sum_{i=1}^{n} \alpha_i y_i z_i$, so y_i , α_i , x_i , and b are all known variables. Therefore, the classification function takes the form of the Equation (8), in which the kernel function uses the Gaussian kernel function $K(x_i, x_j) = \exp(-\frac{||x_i - x_j||^2}{2\sigma^2})$.

$$g(x) = \operatorname{sign}(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b)$$
(8)

Step (3): This paper adopts the one-to-one method [36]. Whenever a new project enters the model, the classifiers 1–15 will respectively give a result. In the end, the category that receives the most votes will be the result of the sample evaluation.

The above is a complete classification model for water quality assessment, but the optimization problem in step (1) still needs to be solved by complicated methods. In 1998, the SMO algorithm that was invented by John Plett solved this problem [37]. This article uses the SMO algorithm to solve the SVM.

In order to solve the optimization problem about $(\alpha_1, \ldots, \alpha_n)$ described by Equation (7), SMO decomposed it into several sub-problems that solve only two parameters each time. The outer loop part traverses the entire set, finds α_i that violates the KKT condition as the first point α_1 , and then performs an inner loop to find the point α_2 that can maximize $|E_1 - E_2|$. Then, update some parameters, find the value of *b*, and finally determine whether the stop condition is met. If it is not satisfied, continue the loop.



Figure 3. Structure of multi-classification model.

3. Improved Membership Function

At present, the basic framework of the water quality evaluation model is already available, but the membership function as the core of the FSVM model has not yet been determined. Because some existing membership functions still have certain defects, this part mainly analyzes and improves the problems of them, and it finally proposes the improved membership function of this paper.

3.1. Membership Function

The membership function itself is a mapping between a sample set and the range (0, 1), which indicates the extent to which a sample belongs to a certain situation. Here are a few common membership functions:

(1) Trigonometric membership function

$$\mu(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x \le b \\ \frac{c-x}{c-b} & b < x \le c \\ 0 & x > c \end{cases}$$
(9)

(2) Trapezoidal membership function

$$\mu(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x \le b \\ 1 & b < x \le c \\ \frac{c-x}{c-b} & c < x \le d \\ 0 & x > d \end{cases}$$
(10)

(3) Distance-based membership function in FSVM

$$\mu(x) = \begin{cases} 1 - \frac{d_{i+}}{r_+ + \delta}, y_i = +1\\ 1 - \frac{d_{i-}}{r_- + \delta}, y_i = -1 \end{cases}$$
(11)

where $d_{i+} = ||x_i - x_+||$, $d_{i-} = ||x_i - x_-||$, $r_+ = \max d_{i+}$, $r_- = \max d_{i-}$, x_+ and x_- are the centers of the positive and negative samples, respectively.

3.2. Basic Form of Membership Function

At present, there are a lot of literatures that have put forward their own ideas on the issue of membership function. In [30], the author used the positive and negative class radius (r_+ , r_-), the distance from the point to the center of the class (d_{i+} , d_{i-}), and the distance from the center of the positive class to the center of the negative class (d_{+-}) to construct the membership function μ_i , as follows:

$$\mu_i = \sqrt{\mu_{i1} \cdot \mu_{i2}} \tag{12}$$

$$\mu_{i1} = \begin{cases} 1 - \frac{1}{1 + (r_{+}^{2} - d_{i+}^{2}) + \delta}, y_{i} = +1\\ 1 - \frac{1}{1 + (r_{-}^{2} - d_{i-}^{2}) + \delta}, y_{i} = -1 \end{cases}$$
(13)

$$\mu_{i2} = \begin{cases} \mu_i^+ = \begin{cases} \frac{\delta + d_{i+}}{r_+}, d_{i+} \le d_{+-} \cdot \varepsilon \\ \delta, & d_{i+} > d_{+-} \cdot \varepsilon \\ \frac{\delta + d_{i-}}{r_-}, d_{i-} \le d_{+-} \cdot \varepsilon \\ \delta, & d_{i-} > d_{+-} \cdot \varepsilon \end{cases}$$
(14)

In [32], the author also proposed two new membership functions. One is a combination of the traditional membership function that is based on distance and compactness. The membership function based on distance is the same as Equation (14), and the membership function based on compactness selects the nearest *p* points around the sample point *x_i*. When all of the *p* samples do not belong to the class of *x_i*, *x_i* is judged as noise, which has no effect on the formation of classification plane, so the value of *e_i* is δ . When all of the *p* samples belong to the class of *x_i*, the value given to *e_i* is $(\sum_{j=1}^{p} \frac{1}{d_{ij}})/p$. When *q* of the *p* samples belong to the class of *x_i*, and the rest do not belong to this class, the $\left|\frac{(\sum_{j=1}^{q} \frac{1}{d_{ij}})}{q} - \frac{(\sum_{j=q+1}^{p} \frac{1}{d_{ij}})}{p-q}\right|$ is used as the value of *e_i*. Finally, $\mu_{i2} = \frac{|e_i|}{\max|e_i|}$, $\mu_i = \mu_{i1} \cdot \mu_{i2}$.

The other one is based on the intra-class hyperplane as defined. It only considers the points inside two intra-class hyperplanes. The membership of the outer points is directly defined as a very small positive number δ . According to the distance d_{i+} or d_{i-} from the point to the hyperplane of this class and the distance *D* between two hyperplanes, the membership of the inner points is defined, as follows.

$$\begin{cases} S_{i+} = \frac{d_{i+} - \min_{1 \le j \le t_+} d_{j+}}{\max_{1 \le j \le t_+} d_{i+} - \min_{1 \le j \le t_+} d_{j+}}, d_{i+} \le \lambda D \\ S_{i+} = \delta, \quad d_{i+} > \lambda D \end{cases}$$
(15)

$$\begin{cases} S_{i-} = \frac{d_{i-} - \min_{1 \le j \le t-} d_{j-}}{\max_{1 \le j \le t-} d_{j+} - \min_{1 \le j \le t-} d_{j-}}, d_{i-} \le \lambda D \\ S_{i-} = \delta , d_{i-} > \lambda D \end{cases}$$
(16)

3.3. Design of the Improved Membership Function

3.3.1. Problems with Existing Membership Functions

(1) Basic architecture issues

In addition to satisfying the requirements of the basic value range (0, 1), the membership function that is required by FSVM is mainly to reflect the requirements of noise reduction and to facilitate the formation of classification planes. Many articles, including [30,31], construct a hypersphere based on the class center, and design a membership function based on the distance from the sample point to the sample center. To some extent, this method relies heavily on the geometry of the sample distribution. For example, in the case of Figure 4 below, the two points A and B contribute the same to the construction of the classification plane, but due to the different distance between the two points and their class centers, the values of membership, as calculated by the above method, based on class center are different.



Figure 4. Traditional center-based hypersphere model.

Therefore, this paper decides to use the idea of intra-class hyperplane to design the membership function. As shown in Figure 5, the class centers $x_+ = \frac{1}{n_+} \sum_{i=1}^{n_+} x_i$ and $x_- = \frac{1}{n_-} \sum_{i=1}^{n_-} x_i$ are first obtained, respectively. Then the two intra-class hyperplanes are constructed by the normal vector $W = x_+ - x_-$.

$$I_{+}: W^{T}(x - x_{+}) = 0$$

$$I_{-}: W^{T}(x - x_{-}) = 0$$
(17)



Figure 5. Inner-class hyperplane model.

Thus, the membership function considers only the set U', which includes the sample points inside the two hyperplanes I_+ and I_- . The sample points outside the hyperplanes are no longer considered, because they do not help in the determination of the optimal hyperplane. The above content is described in Equation (18).

$$S_{i} = \delta, \ x \notin U'$$

$$U'_{+} = \{x | (x - x_{+})(x_{-} - x_{+}) \ge 0, \quad x \in U_{+}\}$$

$$U'_{-} = \{x | (x - x_{-})(x_{+} - x_{-}) \ge 0, \quad x \in U_{-}\}$$

$$U'_{+} \cup U'_{-}$$
(18)

where δ is a very small positive number.

(2) The problem of distance-based membership function

The distance-based membership functions that are designed in [30,32] are inversely proportional to the distance d_i , that is, the closer to the center of the class, the larger the value. This design is mainly based on the idea that the closer the point is to the class center, the more it should belong to this class. However, the point outside the boundary of the classification plane satisfies the condition $|w^Tx + b| > 1$, and its relaxation variable $\varepsilon_i = 0$, so the value of *C* does not affect the classification result. Conversely, such a method makes the S_i of the useless point that is closer to the center of the class bigger than the S_i of the point that may be the support vector. Therefore, this paper adopts the idea of membership function that is based on intra-class hyperplane in [32], and it gives the larger function value to the samples that are closer to the boundary zones between two types.

(3) The problem of compactness-based membership function

In [30], the author used a parameter $\lambda \in (0, 1)$ to solve the problem of noise reduction. When the value of d_i is bigger than the product of λ and the distance D between two hyperplanes, $S_i = \delta$ (δ is a small positive number). However, if the situation in Figure 6 below occurs, that is, the final classification plane is not parallel to the two hyperplanes. Here, d_A or d_B is the distance from the point A or B to the intra-class hyperplane. Assuming that the final λ can make $d_A > \lambda \cdot D$, the noise A can be successfully excluded, but the B point satisfying $d_A = d_B$ should be a support vector. However, the condition of $d_B > \lambda \cdot D$ is also satisfied, so B is also treated as noise. Therefore, there may be some problems with this method.



Figure 6. Problem in the use of the parameter λ .

In [32], the author constructed a membership function based on compactness. When *q* samples of the nearest *p* neighbors of the sample x_i are not in the same class as x_i , the centripetal degree is defined as the former one. However, as shown in Figure 7, if p = 5, and the five white points that are shown in Figure 7e, f are the nearest five neighbors of point A and B, the point A in Figure 7a has a certain effect on the formation of the classification plane, but the point B in Figure 7b is obviously a noise point. But, according to the method of [32], A and B have the same membership.



Figure 7. A situation of membership function based on compactness. (**a**,**b**) Distribution of the sample points for the two cases; (**c**,**d**) Identification of neighbors by a circle; (**e**,**f**) Distribution of neighbors.

3.3.2. Improvement Ideas

For the case where the neighbors of a sample point are all of the same class of this point or the neighbors of a sample point are all the points of the different class, it can be directly determined. However, there are two main situations for areas where the positive and negative sample points are mixed together. One is that the points are close to the junction area of two types, so most of the samples in this area should be given correspondingly large values. The other is that the point is a noise point inside a certain class, but the categories of its neighbors are not all different from it, so this situation is different from the case, where all of the surrounding points are of different class.

At the beginning, this paper uses the category of the nearest neighbor as the criterion for evaluating whether the point in the mixed area is noise. If the class of the sample point closest to it is the same as its class, it is judged to be a useful point. If the class of the sample point closest to it is different from its class, the point is determined to be noise, namely,

- (1) When all of the *p* sample points are not in the class of x_i , x_i is noise, and has no effect on the classification plane formation, so the value of c_i is δ .
- (2) When all the *p* sample points are of the same class as *x_i*, the function is designed according to the degree of compactness of the points around *x_i*, so the value of *c_i* is ∑^p_{i=1} 1/d_{ii}.
- (3) When only *q* samples of the *p* points belong to the same class of x_i , c_i takes the value δ or $\sum_{j=1}^{q} \frac{1}{d_{ij}}$, according to the class of its nearest neighbor.

But, in fact, the counterexample, like Figure 8, can be given. If p = 5, what appears under normal conditions should be similar to the case of Figure 8a,c. Since they are close to the classification plane, the two points A and C do not belong to the cases (1) or (2). Then, according to the class of the nearest neighbor, this method determines that point A is noise, and point C is useful for determining the classification plane.

However, it is inaccurate to base only on whether x_i 's nearest neighbor is of the same or different class as x_i . The reason is that although the model in this article can deal with those noise points, they will also interfere with the judgment of the surrounding points. For example, there is a nearest different-class neighbor next to point B. Point B will be judged as noise by this method, but the actual situation is not the same. Similarly, when the two noise points are very close, this method will instead judge point D as the point that contributes to the classification plane. In both cases, other noise points interfere with the determination of the adjacent sample points.

Thus, the situation similar to that in Figure 8b is first solved here. Although some of the noise isolated from the different-class points is removed by the previous case (1), a different-class point

itself still interferes with the determination of the surrounding points. Therefore, this paper decides to treat the judgment of the previous situation (1) as a priori. After finding out the noise points whose neighbors are all the different-class points, the discrimination of case (2) and case (3) no longer consider these noise points in the situation (1).



Figure 8. Counterexample of the first attempt. The figures (**a**–**d**) are four different cases of surrounding neighbors.

In addition, in order to solve the situation like Figure 8d, we must consider the second influencing factor—the number of same-class and different-class points. When a different-class point is the closest neighbor, and the number of surrounding different-class points is more than that of the same-class points, this point is judged as noise (such as point A). When a different-class point is the closest neighbor, but the number of same-class points is more than that of different-class points, the point is judged to be a useful point (such as point B). When a same-class point is the closest neighbor, and the number of same-class point B). When a same-class point is the closest neighbor, and the number of same-class point is more than that of different-class point is judged to be a useful point (such as point B). When a same-class points, the point is judged to be a useful point (such as point C). When a same-class point is the closest neighbor, but the number of different-class points is more than that of similar points, the point is judged as noise (such as point D).

However, it was found that in the condition from A to D discussed above, when the number of surrounding different-class points is more than the number of same-class points, the point is judged as noise regardless of the category of its closest neighbor. Thus, the result of this judgment is exactly the same as the method that considers only the number of same-class points and different-class points. The category of the nearest neighbor has no meaning. Therefore, it can be determined as directly based on the number of neighbors in the different class and in the same class.

If a counterexample is also given to this classification method, it should be similar to the case of Figure 9b,c under normal circumstances. That is, when there are more neighbors of the same class of this point, it should be a normal point. Similarly, when the number of neighbors that belong to the different class of the point is bigger, it should be the noise.



Figure 9. Counterexample of the second attempt. The figures (**a**–**d**) are four different cases of surrounding neighbors.

However, for example, the number of different-class points in the neighbors of point A in Figure 9a is more than that of the same type, but it is not a noise point. Similarly, the number of same-class points around point D in Figure 9d is more than that of the different class, but it is noise. In fact, this kind of idea that is similar to k-nearest neighbor is acceptable. The most important goal of the function S_{i2} is to find the noise point, so the method is acceptable that identifies the point, the neighbors of which have more points of the different class, as noise. But, this judgment will be limited by the value of the parameter p and the local phenomenon interference.

Therefore, another factor to be considered is added here, which is the number of the points belonging to the same class as the point x_i among the three neighbors other than the p nearest neighbors. This factor can avoid the interference of local phenomena on judgment to a certain extent.

In this case, when the number of same-class points in p neighbors around a point is less than the different-class points, this point should be judged as noise. But, if two or three of the three nearest neighbors, except the p points are belong to the same category, it proves that the above situation is only a partial phenomenon, so the sample point is judged to be a normal point (such as point A). If only one or none of the three neighbors except the p points is in the same class, then the sample point is judged to be noise (such as point B). When there are more same-class neighbors of the p point, if only one or none of the three neighbors except the p points is in the same class, the sample point is judged to be noise (such as point B). When there are more same-class neighbors of the p point, if only one or none of the three neighbors except the p points is in the same class, the sample point is judged to be noise (such as point D). If two or three of the three nearest neighbors except the p points belong to the same category, the sample point is judged to be a normal point (such as point C). The Equation (19) shows this function:

$$c_{i} = \begin{cases} \delta, & t_{i} = -1, \ q - \frac{p}{2} < 0\\ \sum_{j=1}^{q} \frac{1}{d_{ij}}, & t_{i} = 1, \ q - \frac{p}{2} < 0\\ \delta, & t_{i} = -1, \ q - \frac{p}{2} > 0\\ \sum_{j=1}^{q} \frac{1}{d_{ij}}, & t_{i} = 1, \ q - \frac{p}{2} > 0 \end{cases}$$
(19)

where if there are two or three points in the three neighbors belonging to the same class as the point x_i , $t_i = 1$. If there are only one or none of the three points in the three neighbors belonging to the same class as the point x_i , $t_i = -1$.

However, in fact, when $t_i = 1$, regardless of whether the value of $(q - \frac{p}{2})$ is positive or negative, the value of c_i is δ . Similarly, when $t_i = -1$, the value of c_i is $\sum_{j=1}^{q} \frac{1}{d_{ij}}$, regardless of the value of $(q - \frac{p}{2})$. Therefore, the value of the function can be only determined by the positive and negative of t_i , namely,

$$c_{i} = \begin{cases} \delta, & t_{i} = -1\\ \sum_{j=1}^{q} \frac{1}{d_{ij}}, & t_{i} = 1 \end{cases}$$
(20)

However, the discriminant of Equation (20), the condition of the three neighbors, except the p points, is actually based on the number of same-class and different-class points around it in essence. The difference between the two judgement conditions is only the value of p. So the addition of three points is actually meaningless. That is, the problem can be solved by using only the idea of the number of same-class and different-class points around the point.

3.3.3. Improved Membership Function

Therefore, based on the above analysis, the membership function of the paper finally combines both the distance and the compactness. The following Equation (21) is the distance-based membership function of this article that is designed for easy classification:

$$S_{i1} = \begin{cases} 1 - \frac{\max_{1 \le j \le t+} d_{j+} - d_{i+}}{\max_{1 \le j \le t+} d_{j+} + \delta}, y_i > 0\\ \max_{1 \le j \le t+} d_{j-} - d_{i-}\\ 1 - \frac{\max_{1 \le j \le t-} d_{j-} + \delta}{\max_{1 \le j \le t-} d_{j-} + \delta}, y_i < 0 \end{cases}$$
(21)

where *t* – and *t*+ are the total number of positive and negative sample points inside the two hyperplanes, respectively, and δ is a small positive number.

When a point is very close to the intra-class hyperplane, it has no effect on the construction of the classification plane, so its membership degree is infinitely close to zero. As the sample point gets closer to the junction zone of the two types of samples, its contribution to the construction of the classification plane is greater. Therefore, its membership is also greater.

However, only one function in the model is obviously insufficient. For example, when a point satisfying the condition $y_i > 0$ is far away from the positive class hyperplane, it is mixed into the negative class sample points. Obviously, it should be a noise point, but the value of the membership function S_1 above will be large. Therefore, it is necessary to adjust the above function. To this end, this paper designs another membership function to solve the problem of noise and isolated points.

The following is the improved compactness-based membership function that is proposed in this paper. Here, distance from the nearest *p* sample points around the sample point x_i to x_i is d_{i1}, d_{i2}, \cdots , d_{ip} , where *p* is an odd number.

(1) When all the *p* sample points are not in the same class as x_i , x_i is judged as noise and it has no effect on the classification plane formation, namely,

$$c_i = \delta \tag{22}$$

- (2) Reselect *p* sample points around the sample point x_i that are closest to it and do not contain the points of the above case (1). This can effectively avoid the interference of a single noise point in some cases (1) to the judgment of surrounding sample points.
 - a. When all of the *p* sample points at this time are not in the same class as x_i , x_i is judged as noise and it has no effect on the classification plane formation, namely,

$$c_i = \delta \tag{23}$$

b. When all of the *p* sample points are of the same class as x_i , the function is designed according to the compactness of the sample points around x_i , that is, the tighter the sample point, the larger the c_i :

$$c_i = \sum_{j=1}^p \frac{1}{d_{ij}}$$
(24)

c. When *q* points in *p* sample points belong to the same class of x_i , the remaining ones do not belong to this class, the value of c_i is as follows:

$$c_{i} = \begin{cases} \delta, & q - \frac{p}{2} < 0\\ \sum_{j=1}^{q} \frac{1}{d_{ij}}, & q - \frac{p}{2} > 0 \end{cases}$$
(25)

Finally, the formula of the second membership function S_{i2} is presented, as follows:

$$S_{i2} = \frac{c_i}{\max c_i} \tag{26}$$

So far, the design of c_i in the case (1) and the case a of (2) completes the aim of noise reduction. The design of the function value $\sum_{j=1}^{q} \frac{1}{d_{ij}}$ in b of the case (2) excludes the effect of isolated points. When a point is isolated, its c_i will be very small. The c of the case (2) compensates for the loopholes of the previous two cases a and b. It uses the idea of k-nearest neighbor to make S_{i2} better distinguish the effect of the sample points in the junction area by the appropriate parameter p. At the same time, together with case (1), it also solves the problem left by S_{i1} , that is, by the effect of determining the noise, the function S_{i2} negates the membership value in S_{i1} , which is largely due to being far away from the hyperplane in the class.

Since S_{i1} only considers the importance of points near the classification plane, it has no ability to handle noise and isolated points. Therefore, this paper constructs S_{i2} to complete the task of removing noise and isolated points, and it makes up for the defects of S_{i1} . Thus, the final membership function of this paper is determined, as follows:

$$S_{i} = \begin{cases} \delta, & x \notin U' \\ S_{i1} \cdot S_{i2}, & x \in U' \end{cases}$$

$$(27)$$

When the compactness of the neighbors around the sample point is constant, the closer the sample point is to the junction area, the greater its membership degree. When the distance between the sample points and the hyperplane in the class is constant, the bigger the compactness of the neighbors of the sample point, the greater the membership degree. For the method of the combination of S_{i1} and S_{i2} , since the two functions can compensate each other after the addition, this article does not adopt the addition method. When considering that noise should be directly rejected, this paper uses the multiplication method that is uncompensated.

Now, please have a look at the issues discussed before:

- (1) For the case of Figure 4, it is clear that this problem has been solved by using the method of intra-class hyperplane instead of hypersphere. The value of S_{i1} of the new function is the same for A and B in Figure 4, but, depending on the situation of the surrounding points of A and B, different S_{i2} may be given. Finally, the result is the combination of S_{i1} and S_{i2}.
- (2) For the case of Figure 5, the new function judges the sample point based on the number of same-class and different-class points around it, rather than the distance to the intra-class hyperplane.
- (3) For the problem of Figure 6, the improved effect has been shown in the analysis of the situation in Figure 9.

Finally, since the model is to be adopted to multi-class classification, the variables are to be converted. These variables include the class centers $\varphi(x_+)$ and $\varphi(x_-)$, distance ||W|| between two class centers, condition that determines whether it is inside or outside the hyperplane, the distance d_{i+} or d_{i-} between the point and the intra-class hyperplane and the distance d_{ij} between the sample points.

$$\varphi(x_{+}) = \frac{1}{n_{+}} \sum_{i=1}^{n_{+}} \varphi(x_{i}), \ \varphi(x_{-}) = \frac{1}{n_{-}} \sum_{i=1}^{n_{-}} \varphi(x_{i})$$
(28)

$$||W||^{2} = ||\varphi(x_{+}) - \varphi(x_{-})||^{2} = \frac{1}{n_{+}^{2}} \sum_{i=1}^{n_{+}} \sum_{j=1}^{n_{+}} K(x_{i}, x_{j}) + \frac{1}{n_{-}^{2}} \sum_{i=1}^{n_{-}} \sum_{j=1}^{n_{-}} K(x_{i}, x_{j}) - \frac{1}{n_{+} \times n_{-}} \sum_{i=1}^{n_{+}} \sum_{j=1}^{n_{-}} K(x_{i}, x_{j})$$

$$(29)$$

$$(\varphi(x_{-}) - \varphi(x_{+})) \cdot (\varphi(x_{i}) - \varphi(x_{+}))$$

$$= \frac{1}{n_{-}} \sum_{j=1}^{n_{-}} K(x_{i}, x_{j}) + \frac{1}{n_{+}^{2}} \sum_{k=1}^{n_{+}} \sum_{j=1}^{n_{+}} K(x_{k}, x_{j}) - \frac{1}{n_{+}} \sum_{j=1}^{n_{+}} K(x_{i}, x_{j}) - \frac{1}{n_{+} \times n_{-}} \sum_{k=1}^{n_{+}} \sum_{j=1}^{n_{-}} K(x_{k}, x_{j})$$

$$(30)$$

Water 2018, 10, 1303

$$(\varphi(x_{+}) - \varphi(x_{-})) \cdot (\varphi(x_{i}) - \varphi(x_{-}))$$

$$= \frac{1}{n_{+}} \sum_{j=1}^{n_{+}} K(x_{i}, x_{j}) + \frac{1}{n_{-}^{2}} \sum_{k=1}^{n_{-}} \sum_{j=1}^{n_{-}} K(x_{k}, x_{j}) - \frac{1}{n_{-}} \sum_{j=1}^{n_{-}} K(x_{i}, x_{j}) - \frac{1}{n_{+} \times n_{-}} \sum_{k=1}^{n_{+}} \sum_{j=1}^{n_{-}} K(x_{k}, x_{j})$$

$$(31)$$

$$d_{i+} = \frac{\left|W^{T}(\varphi(x_{i}) - \varphi(x_{+}))\right|}{||W||} = \frac{\left|(\varphi(x_{-}) - \varphi(x_{+})) \cdot (\varphi(x_{i}) - \varphi(x_{+}))\right|}{||W||}$$
(32)

$$d_{i-} = \frac{\left|W^{T}(\varphi(x_{i}) - \varphi(x_{-}))\right|}{||W||} = \frac{\left|(\varphi(x_{+}) - \varphi(x_{-})) \cdot (\varphi(x_{i}) - \varphi(x_{-}))\right|}{||W||}$$
(33)

$$d_{ij} = \sqrt{K(x_i, x_i) - 2K(x_i, x_j) + K(x_j, x_j)}$$
(34)

3.4. Data Verification

In order to verify whether the improved membership function that is proposed in this paper can really achieve the expected effect, the following experiments are conducted while using the artificial dataset and the UCI standard data to verify the model performance.

3.4.1. Experiment Based on Artificial Data

In order to test the function's ability to recognize noise and outliers in the case of visualization, the improved membership function is first tested by an artificial data set. Randomly place 90 sample points whose horizontal and vertical coordinates are between 0 and 1 on a two-dimensional plane. Take the data whose x_1 in (0, 0.5) as the negative sample point, and the data whose x_1 in (0.5, 1) as the positive sample point. Then, 10 noise points are randomly placed into the data set. The above data was put into the improved FSVM model. Since the sample set is small, *p* takes 3. Since the construction of S_{i1} is relatively simple, the focus is on the value of the function S_{i2} . The above work is repeated ten times and the results are observed. The specific data distribution of a certain time is shown in Figure 10, and the final results are shown in Table 2.



Figure 10. Distribution of artificial data.

It can be seen that among all of the S_{i2} values, 10 samples have a value of 6.17×10^{-6} , which are all the noise points that are placed previously. Another point (0.36, 0.14) can be seen as an isolated point from the figure below, so its S_{i2} value is only 0.0630. Therefore, it can be seen that the improved FSVM

16 of 23

handles these 11 points that are not useful for classification very well. In the same way, the results of the other several experiments are also the same.

у	<i>x</i> ₁	<i>x</i> ₂	<i>S</i> _{<i>i</i>2}	y	<i>x</i> ₁	<i>x</i> ₂	<i>S</i> _{<i>i</i>2}
-1	0.22	0.80	0.3287	$^{-1}$	0.73	0.90	$6.17 imes10^{-6}$
-1	0.41	0.54	0.3761	-1	0.93	0.83	$6.17 imes 10^{-6}$
-1	0.16	0.98	0.1909	1	0.34	0.07	$6.17 imes10^{-6}$
-1	0.12	0.72	0.2546	1	0.22	0.92	$6.17 imes 10^{-6}$
-1	0.17	0.84	0.3883	1	0.06	0.29	6.17×10^{-6}
-1	0.19	0.43	0.3100	1	0.48	0.31	6.17×10^{-6}
-1	0.27	0.47	0.4214	1	0.63	0.21	0.1526
-1	0.28	0.56	0.6710	1	0.95	0.65	0.3475
-1	0.20	0.27	0.1277	1	0.80	0.07	0.2154
-1	0.20	0.75	0.2858	1	0.75	0.41	0.3068
-1	0.26	0.50	0.3991	1	0.81	0.67	0.3720
-1	0.33	0.65	0.1835	1	0.91	0.93	0.2784
-1	0.36	0.14	0.0630	1	0.77	0.81	0.3665
-1	0.20	0.48	0.3500	1	0.60	0.48	0.3823
-1	0.42	0.36	0.0946	1	0.73	0.76	0.3186
-1	0.07	0.79	0.2842	1	0.71	0.42	0.3982
-1	0.03	0.78	0.2730	1	0.98	0.97	0.1617
-1	0.04	0.67	0.4481	1	0.81	0.99	0.1910
-1	0.08	0.13	0.1211	1	0.85	0.86	0.4601
-1	0.16	0.02	0.1879	1	0.86	0.39	0.4178
-1	0.15	0.56	0.3917	1	0.67	0.45	0.9910
-1	0.01	0.30	0.2041	1	0.78	0.78	0.4035
-1	0.27	0.94	0.2795	1	0.78	0.91	0.2237
-1	0.05	0.98	0.1216	1	0.71	0.60	0.1552
-1	0.07	0.29	0.2036	1	0.92	0.15	0.2456
-1	0.32	0.80	0.2002	1	0.87	0.90	0.6229
-1	0.43	0.90	0.4366	1	0.68	0.45	1.0000
-1	0.43	0.60	0.5239	1	0.73	0.21	0.1997
-1	0.29	0.88	0.2491	1	0.69	0.90	0.1696
-1	0.50	0.94	0.2531	1	0.89	0.76	0.2117
-1	0.28	0.55	0.7064	1	0.87	0.88	0.6807
-1	0.26	0.73	0.2413	1	0.72	0.28	0.2682
-1	0.17	0.58	0.3887	1	0.85	0.67	0.3610
-1	0.22	0.03	0.1858	1	0.97	0.66	0.3351
-1	0.25	0.45	0.3884	1	0.89	0.12	0.2691
-1	0.04	0.65	0.4183	1	0.85	0.41	0.4281
-1	0.44	0.52	0.3003	1	0.69	0.72	0.2458
-1	0.03	0.37	0.1829	1	0.80	0.28	0.2356
-1	0.22	0.94	0.2751	1	0.53	0.83	0.2377
-1	0.41	0.83	0.2080	1	0.61	0.39	0.2323
-1	0.20	0.85	0.4059	1	0.92	0.50	0.1605
-1	0.31	0.37	0.1803	1	0.51	0.86	0.2558
-1	0.41	0.59	0.5391	1	0.54	0.51	0.1625
-1	0.44	0.87	0.4405	1	0.83	0.57	0.1810
-1	0.47	0.93	0.3934	1	0.75	0.33	0.2900
-1	0.10	0.67	0.3253	1	0.61	0.46	0.3996
-1	0.76	0.25	$6.17 imes10^{-6}$	1	0.79	0.71	0.3303
-1	0.58	0.88	$6.17 imes 10^{-6}$	1	0.56	0.88	0.1994
-1	0.85	0.56	$6.17 imes 10^{-6}$	1	0.84	0.72	0.3496
-1	0.60	0.28	$6.17 imes 10^{-6}$	1	0.80	0.02	0.1947

Table 2. Results of experiments on artificial dataset.

3.4.2. Experiment Based on UCI Dataset

Here, the model is validated by three test datasets: column, seeds and haberman. These datasets in the UCI standard databases are characterized by small samples and high-dimensional, which can well verify the performance of improved model of this paper. Firstly, the data is preprocessed and parameter optimized, and the ratio of the training set to the test set in cross-validation is 4:1. The data is put into the standard SVM model, the model based on the intra-class hyperplane in [32], the model based on the intra-class centripetal degree in [32], and the improved FSVM model that is proposed in this paper respectively. The final result that is evaluated by accuracy, recall, and F1-score is shown, as follows:

It can be seen from Table 3 that the distribution of points in the dataset may be too neat, so all of the last three models (FSVM1, FSVM2, and FSVM) have no room for improvement by noise identification. To put it another way, although FSVM cannot further improve the model performance, at least it can prove that FSVM can also handle this standard dataset well. The final performance of FSVM is not weaker than that of standard SVM. It can be seen from Tables 4 and 5 that the F1-score of the improved FSVM method that is proposed in this paper is, respectively, 3% and 6% higher than the standard SVM method. However, the situation of noise may not be complicated, that is, the noise is very easy to be identified. Therefore, FSVM1 and FSVM2 can also achieve the effect of the FSVM model in the seeds dataset. Similarly, FSVM1 can also achieve the effect of the FSVM model in the haberman dataset. However, the above results can only prove that the performance of FSVM in this article is indeed improved when compared to the standard SVM, whether the improvement based on the previous FSVM models is effective cannot be proved here. To this end, other experiments are still needed.

Table 3. Results of experiment on the column dataset.

	SVM	FSVM1	FSVM2	FSVM
Precision	0.8312	0.8312	0.8312	0.8312
Recall	0.7512	0.7512	0.7512	0.7512
F1-score	0.7892	0.7892	0.7892	0.7892

Table 4. Results of experiment on the seeds dataset.

	SVM	FSVM1	FSVM2	FSVM
Precision	0.7934	0.8167	0.8167	0.8167
Recall	0.7619	0.7857	0.7857	0.7857
F1-score	0.7774	0.8009	0.8009	0.8009

Table 5. Results of experiment on the haberman dataset.

	SVM	FSVM1	FSVM2	FSVM
Precision Recall	0.6839 0.5715	0.8750 0.5313	0.5692 0.5882	0.8750 0.5313
F1-score	0.6227	0.6611	0.5785	0.6611

4. Results and Discussion

4.1. Selection of Dataset and Evaluation Indicators

After analyzing the various river basins in China, the paper finally decided to select the water resources monitoring data of the Pearl River Basin for experiments. The Pearl River is China's second largest river. It originates from Maxiong Mountain in the Wumeng Mountains of Yunnan-Kweichow Plateau, flowing through the six provinces of central and western China and northern Vietnam, and finally injects into the South China Sea from the eight inlets downstream. According to data released by the Pearl River Water Resources Bulletin, the total water consumption of the Pearl River in

2016 reached 83.81 billion cubic meters, while the total amount of wastewater discharged was 17.42 billion tons. For this reason, it is of great significance to carry out a water quality assessment on the Pearl River, which is large both in water consumption and waste water discharge [38].

This paper selects the automatic monitoring data of the Pearl River Basin from 2012 to now (6, 2018), a total of 2633 records were deleted after all the null values were removed. Since the pollution of the Pearl River is mainly organic pollution, this paper selects four conventional evaluation indicators: pH value, chemical oxygen demand (CODmn), dissolved oxygen (DO), and ammonia nitrogen (NH₃-N). The distribution of the dataset is very unbalanced as shown in Table 6, and there are many minority classes at the same time, that is, there are several classes at the same time whose number of samples is much smaller than that of the class with the most samples. Therefore, the undersampling method is performed first. The data is then put into the FSVM model based on the improved membership function of this paper.

Value	Count	Percent
Ι	605	22.98%
II	1560	59.25%
III	189	7.18%
IV	103	3.91%
V	95	3.61%
Inferior V	81	3.08%

Table 6. Distribution of water quality data.

4.2. Analysis and Comparison of Evaluation Results

According to the cross-validation method, 80% of the data is used as the training set, and 20% of the data is used as the testing set. The classification is carried out by the FSVM model of this paper, and the results that were obtained are compared with the single factor evaluation (SFE). Finally, part of the different results is shown in Table 7.

No.	pН	DO (mg/L)	CODMn (mg/L)	NH ₃ -N (mg/L)	FSVM	SFE
1	7.76	8.16	1.8	0.22	1	2
2	8.11	9.52	2.3	0.1	1	2
3	8.13	8.92	2.6	0.15	1	2
4	7.92	9.56	1.5	0.2	1	2
5	8.18	8.52	2.1	0.13	1	2
6	7.75	10.7	2.4	0.36	1	2
7	7.95	9.65	2.2	0.2	1	2
8	8.11	8.59	1.6	0.17	1	2
9	7.86	11.1	2.1	0.35	1	2
10	7.61	5.82	4.2	0.14	1	3
11	6.55	5.14	2.8	0.11	1	3

Table 7. Part of the comparison of evaluation results.

The current domestic water quality assessment generally adopts the single factor evaluation method, which is to determine the over-index of each indicator, and the worst one is used as the sample's water quality evaluation result [39]. The advantage of this method is that the calculation is very simple, and because it adopts the principle of pessimism, it is very suitable for the evaluation of drinking water, which attaches great importance to security issues.

However, the negative evaluation results have not allowed the government to grasp the overall situation of water pollution in a certain basin in time. The FSVM model proposed in this paper can solve this problem. The combination of this model and the single-factor evaluation can simultaneously allocate water resources reasonably and provide the comprehensive pollution status of an area.

Most evaluation results of the two models are the same, and the main differences appear in the classes I, II, and III. Part of the water quality standard is shown in Table 1. It is obvious that the dissolved oxygen and chemical oxygen demand of the first and fourth samples meet the Class I standard, but the ammonia nitrogen of the two samples exceeds the Class I standard, so, according to the pessimistic principle, they should be classified as Class II. However, the comprehensive evaluation can classify them as the Class I. Similarly, the chemical oxygen demand of the No.2, No.3 and No.5 samples does not meet the Class I water quality standard, and the other indicators conform to the Class I standard, so the evaluation results are different. Both the chemical oxygen demand and ammonia nitrogen of the samples No.6, No.7, and No.9 exceeded the standard, but their dissolved oxygen was very high. Moreover, the two indicators of these points only exceeded a little, so the comprehensive evaluation result is the first category. It can be seen that it is feasible to use the improved FSVM model to evaluate water quality, and this model can provide comprehensive evaluation results well.

4.3. Analysis and Comparison of Model Performance

The previous data is also put into the standard SVM model and the two methods that are described above based on the intra-class hyperplane and the intra-class centripetal degree method. Since the single factor evaluation can be used to carefully divide water resources, the ultimate goal of the FSVM model is only to provide the manager with real-time overall information.

On the one hand, for the determination of excellent water quality, we hope that the water qualities that are judged to be excellent are really good. That is, each sample that is judged into good-quality class is deserved. Even if the standards are harsh, some of the good water resources cannot be selected. Otherwise, if some water resources that have begun to become polluted are also misjudged as good quality, the managers cannot take immediate measures. Thus, the indicator precision is needed. On the other hand, for polluted areas, we want to identify all of the heavily polluted areas, and do not want some places to "escape" the scope. Similarly, this will also cause delay in remediation. Therefore, the recognition range needs to be widened, even if some of the slightly polluted areas are included in it. Thus, it is necessary to use the indicator recall.

Therefore, we need both of them to achieve the goal of model building, but it is impossible to have the best of both indicators at the same time. If you want to identify more samples of a certain class, there will inevitably be misjudgments. Thus, the paper finally decided to adopt F1-score to consider both the model's precision and recall [40].

The performance of the model is shown in Table 8. It can be clearly seen that the improved FSVM model is better than several other models, and its F1-score is 8% higher than the standard SVM model. All of these show that the model in this paper can better evaluate the water quality comprehensively. The reason for the poor performance of the FSVM2 model here might be that it is specifically designed for gene classification problems.

	SVM	FSVM1	FSVM2	FSVM
Precision	0.7609	0.8345	0.5840	0.8395
Recall	0.6979	0.7083	0.6563	0.7396
F1-score	0.7280	0.7663	0.6180	0.7864

Table 8. Evaluation index of these models.

5. Conclusions

Water is closely related to people's lives, and it is an indispensable resource. However, with the deterioration of industrial pollution, it is imperative to classify water resources in different regions in order to use water more reasonably, efficiently, and safely. For the consideration of people's safety, the current domestic evaluation method of water quality is single factor assessment. This method can eliminate the potential water misuse through negative evaluation. However, at the same time, because the result of the evaluation is too pessimistic, it will not be able to provide the current situation of water

pollution in a certain area from a global perspective. Thus, timely measures cannot be implemented immediately. Therefore, a model that can comprehensively evaluate water quality is essential.

After understanding the traditional assessment methods of water quality and some evaluation methods in recent years, this paper finally decided to build a classification model that is based on the support vector machine. Firstly, the model is optimized by data preprocessing, data balance, cross-validation and parameter optimization. Then, since the samples do not fully comply with the water quality standards set by the state in many cases, noise is likely to occur in some ambiguous areas. Therefore, the standard SVM model is optimized in terms of removing the influence of noise and isolated points, so the membership function is adopted to form the FSVM model.

However, the membership function proposed in several papers has certain problems in some cases. For this reason, this paper builds a new membership function step by step. The distance-based function and the compactness-based function have been improved successively. The closer the point is to the classification plane, the larger the distance-based function. The compactness-based function first discriminates the influence of a part of the noise points by a priori, and it then determines whether a sample point is noise by the number of same-class and different-class points in the surrounding p neighbors. Finally, the two functions are combined into the new membership function.

In order to verify whether the improved membership function of this paper is reasonable in practical applications, three experiments are carried out in this paper. Firstly, an experiment is done through an artificial dataset, which is intended to observe whether the function is meaningful through the two-dimensional visual data. Then, through the high-dimensional data in the UCI database, the performance of the function is tested. However, the above two kinds of data are not the actual data of water quality assessment. Therefore, the water quality monitoring historical data is finally used for the experiment.

Finally, the result of the experiment on an artificial dataset shows that the model can deal with the negative effects of noise and isolated points. The result of the experiment on the UCI dataset shows that the model does have good performance when dealing with the high-dimensional actual data. The result of the experiment on water quality monitoring historical data shows that the improved FSVM model proposed in this paper is indeed better than the previous models to some extent. In the field of water quality assessment, it can complete the comprehensive evaluation task well, and it can provide the overall information better.

Author Contributions: Conceptualization, W.S. and S.C.; Data curation, S.C.; Formal analysis, W.S. and S.C.; Funding acquisition, W.S.; Investigation, S.C.; Methodology, S.C. and C.L.; Project administration, W.S.; Resources, W.S.; Software, S.C. and C.L.; Supervision, W.S.; Validation, S.C.; Visualization, S.C.; Writing—original draft, W.S. and S.C.; Writing—review & editing, W.S., S.C. and C.L.

Funding: This work was supported by the National Natural Science Foundation of China (No. 71371025), and the Beijing Natural Science Foundation (No. 9182010).

Acknowledgments: The authors are very grateful for the insightful comments and suggestions of the anonymous reviewers and the editor, which have helped to significantly improve this article.

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

References

- 1. Galaitsi, S.E.; Russell, R.; Bishara, A.; Durant, J.L.; Bogle, J.; Huber-Lee, A. Intermittent Domestic Water Supply: A Critical Review and Analysis of Causal-Consequential Pathways. *Water* **2016**, *8*, 274. [CrossRef]
- 2. Kou, L.; Li, X.; Lin, J.; Kang, J. Simulation of Urban Water Resources in Xiamen Based on a WEAP Model. *Water* **2018**, *10*, 732. [CrossRef]
- 3. Huang, X.; Chen, X.; Huang, P. Research on Fuzzy Cooperative Game Model of Allocation of Pollution Discharge Rights. *Water* **2018**, *10*, 662. [CrossRef]
- 4. Liu, Y.; Zhang, J.; Zhao, Y. The Risk Assessment of River Water Pollution Based on a Modified Non-Linear Model. *Water* **2018**, *10*, 362. [CrossRef]

- 5. Liyanage, C.P.; Yamada, K. Impact of population growth on the water quality of natural water bodies. *Sustainability* **2017**, *9*, 1405. [CrossRef]
- 6. Guo, J.S.; Wang, H.; Long, T.Y. Analysis and Development of Water Quality Evaluation Method. *Chongqing Environ. Sci.* **1999**, *21*, 1–3.
- 7. Tao, T.; Sun, S.; Jiang, D.; Fang, H. Fuzzy Comprehensive Evaluation Apply in Water Quality Assessment of Chaohu Lake. *Environ. Sci. Manag.* **2010**, *35*, 177–180.
- 8. Gao, J. Application of Fuzzy Comprehensive Evalution to Water Quality of Binhe Park. *J. Taiyuan Normal Univ.* **2011**, *3*, 106–109.
- Ding, X.; Chong, X.; Bao, Z.; Xue, Y.; Zhang, S. Fuzzy Comprehensive Assessment Method Based on the Entropy Weight Method and Its Application in the Water Environmental Safety Evaluation of the Heshangshan Drinking Water Source Area, Three Gorges Reservoir Area, China. *Water* 2017, *9*, 329. [CrossRef]
- 10. Wang, H.M.; Lu, W.X.; Xin, G.; Wang, H.X. Application of Grey Clustering Method for Surface Water Quality Evaluation. *Water Sav. Irrig.* **2007**, *5*, 20–22.
- 11. Deng, X.; Chen, Q.; Zhang, J. Application of Grey Clustering Method to Water Quality Evaluation in Jinjiang River of Fujian Province. *Environ. Sci. Manag.* **2010**, *35*, 187–191.
- Yun, Y.; Zou, Z. An improved synthetic evaluation method on water quality evaluation in city sections of the Three Gorges reservoir area. In Proceedings of the IEEE International Conference on Grey Systems and Intelligent Services, Nanjing, China, 18–20 November 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 289–293.
- Li, Y.; Zhou, J.; Wang, X.; Zhou, X. Water Quality Evaluation of Nearshore Area Using Artificial Neural Network Model. In Proceedings of the International Conference on Bioinformatics and Biomedical Engineering, Beijing, China, 11–13 June 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–4.
- Li, S.; Zhao, N.; Shi, Z.; Tang, F. Application of artificial neural network on water quality evaluation of Fuyang River in Handan city. In Proceedings of the International Conference on Mechanic Automation and Control Engineering, Wuhan, China, 26–28 June 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1829–1832.
- 15. Hua, Z.L.; Qian, W.; Li, G.U. Application of improved LM-BP neural network in water quality evaluation. *Water Resour. Prot.* **2008**, *24*, 22–25.
- 16. Liu, K. Fuzzy Probabilistic Neural Network Water Quality Evaluation Model and Its Application. *J. China Hydrol.* **2007**, *1*, 31, 42–45.
- 17. Yang, Z.M. Uncertainty Support Vector Machine; Science Press: Beijing, China, 2012.
- Deng, N.; Tian, Y.; Zhang, C. Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions; Chapman & Hall/CRC: Boca Raton, FL, USA, 2012.
- Samadzadegan, F.; Hasani, H.; Schenk, T. Simultaneous feature selection and SVM parameter determination in classification of hyperspectral imagery using Ant Colony Optimization. *Can. J. Remote Sens.* 2012, *38*, 139–156. [CrossRef]
- 20. Syarif, I.; Prugel-Bennett, A.; Wills, G. SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance. *Telkomnika* **2016**, *14*, 1502–1509. [CrossRef]
- 21. Huang, C.L.; Dun, J.F. A distributed PSO–SVM hybrid system with feature selection and parameter optimization. *Appl. Soft Comput. J.* 2008, *8*, 1381–1391. [CrossRef]
- 22. Fei, B.; Liu, J. Binary tree of SVM: A new fast multiclass training and classification algorithm. *IEEE Trans. Neural Netw.* **2006**, *17*, 696–704. [CrossRef] [PubMed]
- 23. Liu, C.; Wang, W.; Wang, M.; Lv, F.; Konan, M. An efficient instance selection algorithm to reconstruct training set for support vector machine. *Knowl.-Based Syst.* **2016**, *116*, 58–73. [CrossRef]
- 24. Li, X.; Wang, L.; Sung, E. AdaBoost with SVM-based component classifiers. *Eng. Appl. Artif. Intell.* 2008, 21, 785–795. [CrossRef]
- 25. Suykens, J.A.K.; Vandewalle, J. *Least Squares Support Vector Machine Classifiers*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1999.
- 26. Zhou, Z.Y.; Wang, X.L. Water quality evaluation based on Support Vector Machine with parameters optimized by genetic algorithm. *Comput. Eng. Appl.* **2008**, *44*, 190–193.
- 27. Chen, L.; Liu, J.M.; Liu, X.X. Application of Support Vector Machine in the groundwater quality evaluation. *J. Northwest A F Univ.* **2010**, *38*, 221–226.
- 28. Dai, H.L. Forecasting and evaluating water quality of Changjiang River based on composite least square SVM with intelligent genetic algorithms. *Appl. Res. Comput.* **2009**, *26*, 79–81.

- 29. Lin, C.F.; Wang, S.D. Fuzzy support vector machines. IEEE Trans. Neural Netw. 2002, 13, 464-471. [PubMed]
- 30. Ren, Y.F. Some Studies of SVM Model Improvement. Master's Thesis, Nanjing University of Posts and Telecommunications, Nanjing, China, 2013.
- 31. Wu, M. Some Researches on the Algorithm of Support Vector Machine Classification. Master's Thesis, Nanjing University of Posts and Telecommunications, Nanjing, China, 2014.
- 32. Xu, C.Y. Research of Fuzzy Support Vector Machine and its Application of Gene Classification. Master's Thesis, Nanjing Forestry University, Nanjing, China, 2013.
- 33. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. J. Artif. Intell. Res. 2002, 16, 321–357. [CrossRef]
- Hu, S.; Liang, Y.; Ma, L.; He, Y. MSMOTE: Improving Classification Performance When Training Data is Imbalanced. In Proceedings of the International Workshop on Computer Science & Engineering, Qingdao, China, 28–30 October 2009; pp. 13–17.
- 35. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; Morgan Kaufmann Publishers Inc.: Burlington, MA, USA, 1995; pp. 1137–1143.
- 36. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [PubMed]
- 37. Platt, J.C. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. In *Advances in Kernel Methods-Support Vector Learning*; Philomel Books: New York, NY, USA, 1999; pp. 212–223.
- 38. Liu, L.; Jiang, T.; Xu, H.; Wang, Y. Potential Threats from Variations of Hydrological Parameters to the Yellow River and Pearl River Basins in China over the Next 30 Years. *Water* **2018**, *10*, 883. [CrossRef]
- Xu, Z.X. Single Factor Water Quality Identification Index for Environmental Quality Assessment of Surface Water. J. Tongji Univ. 2005, 33, 482–488.
- Goutte, C.; Gaussier, E. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In Proceedings of the European Conference on Information Retrieval, Santiago de Compostela, Spain, 21–23 March 2015; Springer: Berlin/Heidelberg, Germany, 2005; pp. 345–359.



© 2018 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/).