

Article

Outlier Detection Using Improved Support Vector Data Description in Wireless Sensor Networks

Pei Shi ^{1,2} , Guanghui Li ^{1,*} , Yongming Yuan ² and Liang Kuang ³¹ School of IoT Engineering, Jiangnan University, Wuxi 214122, China; ship@ffrc.cn² Freshwater Fisheries Research Center of Chinese Academy of Fishery Sciences, Wuxi 214081, China; Yuan@ffrc.cn³ School of IoT Engineering, Jiangsu Vocational College of Information Technology, Wuxi 214153, China; kuangliang.89@163.com

* Correspondence: ghli@jiangnan.edu.cn; Tel.: +86-510-85197572

Received: 23 September 2019; Accepted: 27 October 2019; Published: 30 October 2019



Abstract: Wireless sensor networks (WSNs) are susceptible to faults in sensor data. Outlier detection is crucial for ensuring the quality of data analysis in WSNs. This paper proposes a novel improved support vector data description method (ID-SVDD) to effectively detect outliers of sensor data. ID-SVDD utilizes the density distribution of data to compensate SVDD. The Parzen-window algorithm is applied to calculate the relative density for each data point in a data set. Meanwhile, we use Mahalanobis distance (MD) to improve the Gaussian function in Parzen-window density estimation. Through combining new relative density weight with SVDD, this approach can efficiently map the data points from sparse space to high-density space. In order to assess the outlier detection performance, the ID-SVDD algorithm was implemented on several datasets. The experimental results demonstrated that ID-SVDD achieved high performance, and could be applied in real water quality monitoring.

Keywords: wireless sensor networks (WSNs); outlier detection; support vector domain description; Parzen-window algorithm; water quality monitoring

1. Introduction

Wireless sensor networks (WSNs) have been widely used in various fields, such as industrial (e.g., industrial surveillance), military (e.g., military reconnaissance), medical (e.g., medical diagnosis), agricultural (e.g., agriculture production detection), mechanical engineering, and aerospace engineering applications [1–8]. The reliability of sensor data has increasingly attracted attention from both academia and industry. Outlier detection can recognize noise, errors, events, and hostile attacks, which helps to reduce network risk and ensure data quality [9–12]. Generally, outliers are less than the normal data in the monitoring process, and they can represent changes in monitoring objects and environments. Therefore, outliers have great potential value. In aquaculture, sensors are susceptible to germ corrosion and easily go wrong since they are deployed underwater [13]. Moreover, in the water quality monitoring process, a high speed of outlier detection is required for processing big data.

There are four commonly used outlier detection methods: statistical-based [14,15], nearest neighbor [16,17], clustering-based [18,19], classification-based [20–22], etc. However, these methods still have some limitations when they are used in practice. The statistical-based methods construct models based on prior knowledge [23]. There is no mathematical model to match the real application problem of WSNs perfectly. The nearest neighbor method is a classic detection algorithm [24], and it is time-consuming and has poor scalability when applied in high-dimensional data. Clustering-based methods are limited to the issue of clustering width [25]. Meanwhile, the calculation of data distance

consumes computational resources in high-dimensional datasets. Therefore, this method is unsuitable for the limited-power devices used in WSN applications. Classification-based methods include Bayesian network theory, support vector machine, etc. Bayesian networks can obtain the correlations among data, but they have poor scalability for high-dimensional data. Although the calculation is complex, support vector machines (SVMs) have been introduced to outlier detection for their advantages in solving binary classification problems.

Support vector data description (SVDD) is a widely used one-class support vector machine (OC-SVM) proposed by Tax and Duin [26]. It is an unsupervised learning method suitable for detecting outliers in the fault monitoring process. Jae [27] applied SVDD to classify normal behavior patterns and to detect abnormal behavioral patterns. Bovolo [28] utilized change vector analysis (CVA) and SVDD for a change detection problem. Khediri [29] presented a procedure based on the kernel k -means clustering and SVDD to separate different nonlinear process modes and to effectively detect faults in the etch metal process. Liu [30] presented a high-speed inline defect inspection scheme based on fast SVDD for a thin-film transistor (TFT) array process of thin-film transistor liquid crystal display (TFT-LCD) manufacturing. Zhao [31] used an SVDD-based method in pattern-recognition-based chiller fault detection.

However, most outlier detection algorithms based on SVDD only take the kernel-based distance between the spherical boundary and the data point into account, ignoring the distribution of data. Therefore, many researchers have conducted experiments on SVDD to improve the fault detection performance [32,33]. Lee [32] proposed a distance measurement to SVDD based on the notion of a relative density degree for each data point in order to reflect the distribution of a given dataset. Cha [33] imported the notion of density weight to SVDD, which is the relative density of each data point based on the density distribution of the target data using the k -nearest neighbor (k -NN) approach. However, this approach requires more calculation. When used in an unbalanced dataset, the performance will be unstable.

In this paper, we develop a new method based on existing studies [32,33], namely, the improved density-compensated SVDD algorithm (ID-SVDD). The relative density weight is used to search for an optimal SVDD. In contrast to the existing studies, we obtain the relative density weight by the exponentially weighted Parzen-window density. We incorporate it with SVDD to help obtain the distribution of the target data. All data points are efficiently mapped from sparse space to high-density space. Then, the Mahalanobis distance is utilized to improve the Gaussian window function, which can eliminate the interference of correlations between variables. The traditional SVDD, density-weighted SVDD (DW-SVDD), and density-compensated SVDD (D-SVDD) are compared with ID-SVDD. Experimental results indicate that the detection accuracy and efficiency were both improved by ID-SVDD, and that it could be applied for outlier detection in real water quality monitoring processes.

The paper is organized as follows: in Section 2, we introduce the traditional SVDD method and the ID-SVDD method. In Section 3, we conduct experiments and demonstrate the effectiveness of ID-SVDD for outlier detection. In Section 4, we give conclusions for this study.

2. Methodology

2.1. Support Vector Data Description

SVDD is a common data classification algorithm proposed by Sun and Tsung [34]. The basic idea is to map target data to high-dimensional feature space, and constructing a data description as the smallest sphere to contain all possible target data [35]. The objective of SVDD is to find a spherical boundary with minimal radius R with center o , and realize the classification of unknown data. The data points inside the sphere are the target class, and those outside are treated as the non-target class. For ease of reference, Table 1 summarizes the key notations. The details of SVDD are illustrated in Figure 1.

Table 1. Key notations.

Symbol	Description
R	Radius of sphere
o	Center of sphere
C	The trade-off between sphere volume and the number of target data outside the sphere
ξ_i	Slack variable
α	Lagrange multiplier
R	The distance between an observation datum in the feature space and center a
θ	The mean of Parzen-window density $Par(x_i)$
d	The feature dimension of input data
w	Weighting factor
n	The number of target data
ρ_i	Relative density weight of x_i
$Par(x_i)$	Parzen-window density of x_i
md_{ij}	Mahalanobis distance between vectors
MS	Covariance matrix
\bar{x}	Mean value of x_i
P	Relative density weight array
TP	The number of true positive results
TN	The number of true negative results
FP	The number of false positive results
FN	The number of false negative results
m	The degree of polynomia
δ	Bandwidth of Gaussian kernel function
k	A constant
e	A constant

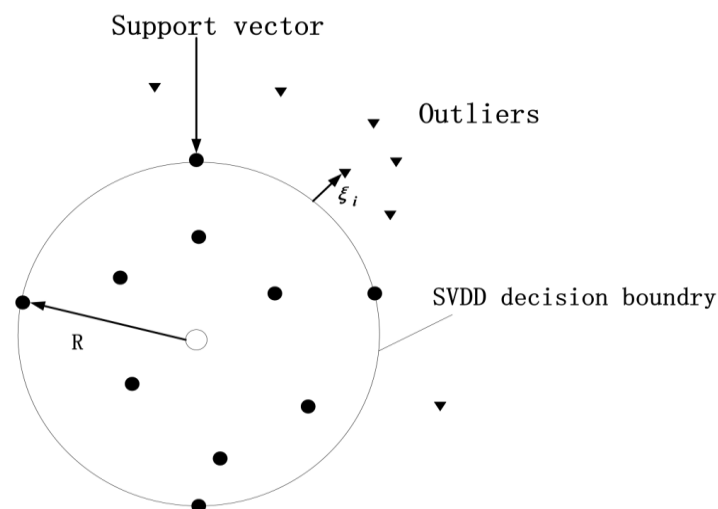


Figure 1. Illustration of support vector data description (SVDD) in feature space for outlier detection.

Given target data $\{x_i, i = 1, 2, \dots, n\}$, SVDD maps the target data from input space into a feature space F via nonlinear mapping function φ and gets the smallest sphere $\Omega = (o, R)$ in F . The objective function of SVDD is as follows:

$$\begin{cases} \min F(R, a, \xi_i) = R^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t. } \|\varphi(x_i) - a\|^2 \leq R^2 + \xi_i, \xi_i \geq 0, \forall i = 1, 2, \dots, n \end{cases} \quad (1)$$

where C is a parameter that denotes the trade-off between sphere volume and the number of target data outside the sphere. The slack variable ξ_i is used to incorporate the effect of data not included in the data description, which allows a probability that some points can be wrongly classified.

To solve the objective function (1), we introduce the Lagrange multiplier α . By calculating the inner product with the kernel function, we can get the rearranged function as follows:

$$\begin{cases} \max \sum_{i=1}^n \alpha_i K(x_i \cdot x_i) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i \cdot x_j) \\ \text{s.t. } 0 \leq \alpha_i \leq C, i = 1, 2, \dots, n \\ \sum_{i=1}^n \alpha_i = 1 \end{cases} \quad (2)$$

where $K(x_i, x_j)$ is the kernel function that satisfies Mercer's theorem [36]. The radius R of the sphere and the distance r between an observation datum in the feature space and center o are denoted as:

$$R^2 = K(z \cdot z) - 2 \sum_{i=1}^n \alpha_i K(z \cdot x_i) + \sum_{i=1, j=1}^n \alpha_i \alpha_j K(x_i \cdot x_j) \quad (3)$$

$$r^2 = K(x_k \cdot x_k) - 2 \sum_{i=1}^n \alpha_i K(x_i \cdot x_k) + \sum_{i=1, j=1}^n \alpha_i \alpha_j K(x_i \cdot x_j) \quad (4)$$

In outlier detection, Figure 1 also shows the SVDD detection principle. It determines the description boundary as the detection boundary. For given test data, x is regarded as a target datum inside the sphere if $r \leq R$, which indicates x is normal. Otherwise, it is treated as an outlier, which indicates x is abnormal.

2.2. Density-Compensated Support Vector Data Description

The traditional SVDD algorithm often ignores the impact of data density distribution on classification [33], which means the sphere cannot reflect all features of the target data, and reduces the classification accuracy. To account for the distribution information of data, we introduce the notion of relative density weight to compensate SVDD, which reflects how dense the region of target data is compared to other regions. This approach makes the training data in high-density areas more likely to fall into the sphere than those in low-density areas.

In this paper, the Parzen-window algorithm [37] is applied to calculate the relative density weight of sample data. Assuming the target data $X = \{x_1, x_2, \dots, x_i, i = 1, 2, \dots, n\}$, the relative density weight of point x_i in a dataset is determined as:

$$\rho_i = \exp \left\{ \omega \times \frac{Par(x_i)}{\theta} \right\}, \forall i = 1, 2, \dots, n \quad (5)$$

$$Par(x_i) = \frac{1}{n} \sum_{j=1}^n \left(\frac{1}{\sqrt{(2\pi)^{d_s}}} \right) \exp \left(-\frac{1}{2s} \|x_i - x_j\|^2 \right) \quad (6)$$

where ρ_i is the relative density weight, $\theta = \frac{1}{n} \sum_{i=1}^n Par(x_i)$ is the mean Parzen-window density $Par(x_i)$, d represents the feature dimension of input data, ω ($0 \leq \omega \leq 1$) is the weighting factor, s denotes the smoothing parameter of Parzen-window density, and n is the number of target data.

We use relative density to reflect the data distribution in real space. In the process of searching for an appropriate description, we calculate the relative density weight according to Equation (5). After importing the relative density weight to SVDD, we obtain the redefined objective function as follows:

$$\begin{cases} \min F(R, a, \xi_i) = R^2 + C\rho(x_i) \sum_{i=1}^n \xi_i \\ \text{s.t. } \|\varphi(x_i) - a\|^2 \leq R^2 + \xi_i, \xi_i \geq 0, \forall i = 1, 2, \dots, n \end{cases} \quad (7)$$

Let the relative density weight multiply the slack variable. Then each datum in high-density regions will get a high relative density value. For searching the optimal description of target data, D-SVDD can shift the description boundary to the dense areas. By introducing the Lagrange multiplier to solve Equation (7) in SVDD, we get the optimization Equation (8):

$$\begin{cases} \max \sum_{i=1}^n \alpha_i K(x_i \cdot x_i) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j K(x_i \cdot x_j) \\ \text{s.t. } 0 \leq \alpha_i \leq \rho(x_i) C, i = 1, 2, \dots, n \\ \sum_{i=1}^n \alpha_i = 1 \end{cases} \quad (8)$$

2.3. Outlier Detection Using Improved Density-Compensated SVDD

In D-SVDD, we choose the Gaussian function as the window function of the Parzen-window algorithm and use the Euclidean distance to measure the distance in the Gaussian function. However, Euclidean distance does not take into account the correlation between sample points [38], which will affect the precision of D-SVDD. Mahalanobis distance (MD) is scale-invariant [39], which can overcome the shortcomings of Euclidean distance. Thus, it can avoid the calculation error caused by measurement units or the difference in magnitude of eigenvector values [40]. The performance of MD is better than Euclidean distance. MD is a non-uniform distribution of the normalized distance in Euclidean space, and it is constant for all linear transformations. The formula of MD is given as follows:

$$md_{ij} = \sqrt{(x_i - x_j)^T S^{-1} (x_i - x_j)} \quad (9)$$

$$MS = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad (10)$$

where md_{ij} is the Mahalanobis distance between vector x_i and x_j , MS is the covariance matrix between two vectors, and $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ denotes the mean of x_i .

In this paper, we introduce the MD to replace Euclidean distance in the Gaussian function. By calculating the MD between data points, the improved Parzen-window density for x_1 is redefined. The Parzen-window relative density weight is denoted as

$$IPar(x_i) = \frac{1}{n} \sum_{j=1}^n \left(\frac{1}{\sqrt{(2\pi)^{d_s}}} \right) \exp\left(-\frac{1}{2s} md_{ij}^2\right) \quad (11)$$

$$\rho_i = \exp\left\{ \omega \times \frac{IPar(x_i)}{\theta} \right\}, \forall i = 1, 2, \dots, n \quad (12)$$

Now, we can give the pseudocode of the ID-SVDD algorithm as shown in Algorithm 1. The outputs of the ID-SVDD algorithm are the Lagrange multipliers α_i , radius R of sphere, and distance r . For a given x_i , it is classified as an outlier if the distance r_i is greater than R . If not, x_i is classified as a normal datum.

Algorithm 1. ID-SVDD outlier detection

Input: Target dataset $X = \{x_1, x_2, \dots, x_i, i = 1, 2, \dots, n\}$, kernel function $K(\cdot)$
Output: α_i , R , and r .
Begin
 Define an array P to store relative density weight for each point.
 for ($k = 1; k \leq n; k++$) do
 calculate $P_k = \rho(x_k)$ according to Equation (12)
 End
 Solve the optimization problem of (8).
 Determine a sample whose α_i is between 0 and $\rho(x_i)C$.
 Calculate the radius R of sphere and the distance r according to Equations (3) and (4).
End
Return α_i , R and r .

3. Experiments

3.1. Experiment Design

In order to evaluate the performance of ID-SVDD, we compared it with the traditional SVDD, D-SVDD, and DW-SVDD provided in [33]. Cha proposed the DW-SVDD with a weight coefficient calculated by k -NN distance. We chose DW-SVDD for comparison because it also attempts to apply the relative density to traditional SVDD. Meanwhile, these four methods were implemented with MATLAB language and run on a PC with 2.9-GHz Core™ processor, 16.0 G memory, and Microsoft Windows 10 operating system.

Considering the completeness and continuity of data, we chose the data of nodes 12 and 17 in the SensorScope system dataset [41] to complete simulation experiments. The SensorScope system is deployed at Grand-St-Bernard Mountain, which lies between Switzerland and Italy. The datasets have two attributes, including the external temperature and surface temperature. In addition, we finished experiments on a real water quality dataset with three attributes, including dissolved oxygen (DO), pH, and dissolved oxygen relative saturation. More detailed information about the experimental datasets is presented in Table 2.

Table 2. Experimental datasets.

Datasets	Attributes	Normal Data	Outliers
SensorScope node12	2	1411	44
SensorScope node17	2	1309	137
water quality data	3	1706	50

We used different indexes to evaluate the performance of ID-SVDD. These indexes included true positive rate (TPR), true negative rate (TNR), accuracy, and run time [42,43]. The calculation formulas of these indexes are shown as follows.

$$TPR = \frac{TP}{TP + FN} \quad (13)$$

$$TNR = \frac{TN}{FP + TN} \quad (14)$$

$$Accuracy = \frac{TP + TN}{(TP + FN) + (FP + TN)} \quad (15)$$

where TP is the number of true positive results, TN represents the number of true negative results, FP is the number of false positive results, FN denotes the number of false negative results. These indicators together with run time can measure the performance of outlier detection methods effectively.

3.2. Experiment Results

3.2.1. Comparison among Different Kernel Functions

The kernel function of SVDD can map nonlinear relations to higher-dimensional space and construct linear regression for processing [44]. In ID-SVDD, the kernel function also plays a key role. The common kernel functions include the linear kernel function (16), polynomial kernel function (17), Gaussian kernel function (18), and Sigmoid kernel function (19) [45]:

$$k(x, z) = x \cdot z \quad (16)$$

$$k(x, z) = ((x \cdot z) + 1)^m \quad (17)$$

$$k(x, z) = \exp\left\{-\frac{\|x - z\|^2}{2\delta^2}\right\} \quad (18)$$

$$k(x, z) = \tanh(k(x, z) + e), (k > 0, e < 0) \quad (19)$$

Here, we set the parameter C within the range 2^{-8} to 2^8 , which controls the trade-off between volume and errors. For better outlier detection results, we conducted experiments to choose the optimal kernel function. For the variable δ in Gaussian kernel function, we set it between 2^{-8} and 2^8 . Further, we used fivefold cross validation (CV) to find the adequate parameters of these kernel functions. After parameter selection, we obtained the optimal results of ID-SVDD with the SensorScope dataset. The detailed results are provided in Table 3.

Table 3. Comparison among different kernel functions. TNR: true negative rate; TPR: true positive rate.

Different Kernel Functions	TPR (%)	TNR (%)	Accuracy (%)
SensorScope12			
Linear	89.3525	0	84.4898
Ploy	100	0	94.5578
Gaussian	99.4245	87.5	98.7755
Tanh	98.1295	20	93.8776
SensorScope17			
Linear	38.4095	28.1482	36.5014
Ploy	92.5550	45.9259	83.8843
Gaussian	100	97.037	99.449
Tanh	68.1895	0	55.5096

Table 3 clearly indicates that the TPR, TNR, and accuracy of Gaussian kernel function were superior to the other three kernel functions in nodes 12 and 17 of SensorScope. Based on these experimental results, we adopted the Gaussian function as the kernel function of ID-SVDD in the outlier detection of water quality data.

Figure 2a,b are the testing results distribution diagrams of the ID-SVDD detection algorithm in SensorScope node 12 and 17 datasets, respectively. From Figure 2, the support vector constructed the boundary to distinguish the normal data from outliers. The decision boundaries of the two datasets are both irregular graphics. The blue points outside the sphere represent the outliers, whereas the red points inside the sphere are normal data. It is clear that the detection model could describe the data edges accurately. So, the ID-SVDD model is an effective detection model.

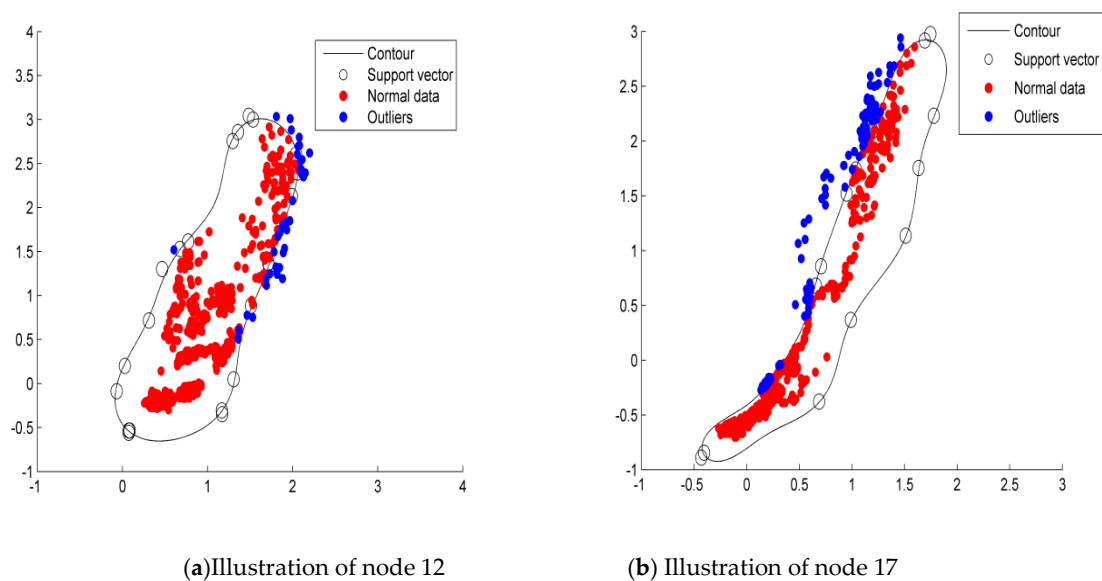


Figure 2. Distribution of the SensorScope dataset.

3.2.2. Comparison Results of Different Datasets

We conducted experiments to compare ID-SVDD with traditional SVDD, D-SVDD, and DW-SVDD in a set of standard datasets from SensorScope. The detection results are displayed in Table 4.

Table 4. Detection results of SensorScope datasets.

	ID-SVDD	D-SVDD	DW-SVDD	SVDD
Node 12				
TPR (%)	99.4245	98.4173	70.5036	98.0496
TNR (%)	87.5	100	82.5	100
Accuracy (%)	98.7755	98.5034	71.1565	98.1788
Time (s)	0.489	0.5329	0.4211	0.5463
Node 17				
TPR (%)	100	98.8338	90.3553	99.3232
TNR (%)	97.037	100	63.7037	98.5185
Accuracy (%)	99.449	98.8981	85.3994	99.1736
Time (s)	0.3794	0.578	0.4172	0.3763

We can see from Table 4 that the TPR and accuracy values of ID-SVDD in nodes 12 and 17 were both superior to the D-SVDD, DW-SVDD, and SVDD. However, the TNR of ID-SVDD was lower than the TNR for D-SVDD and SVDD. These results indicate that the MD improved Parzen-window relative density weight could eliminate the interference of correlation between variables. It is appropriate for measuring the distance between target data. Meanwhile, the TPR, TNR, and accuracy of D-SVDD in nodes 12 and 17 were superior to those for DW-SVDD and SVDD. These results indicate that Parzen-window relative density weight is appropriate for compensating the SVDD. In terms of run time, these four algorithms were close. Actually, the use of ID-SVDD provided an acceptable improvement in outlier detection on the SensorScope datasets of nodes 12 and 17.

3.2.3. Experimental Results on Water Quality Datasets

This experiment evaluated the ID-SVDD algorithm on a real water quality dataset. All data were collected from the internet of things (IOT) monitoring system running in the Nanquan breeding base located in Wuxi city, Jiangsu province [46]. This system uses various types of sensors to collect water quality data (e.g., DO, pH, and dissolved oxygen relative saturation). These data were transmitted from sensors to a server via the IOT monitoring system.

The water quality dataset in this experiment included 1756 data (sampled 10 min once) from 20 May to 2 June 2017. We chose the first 1052 data as a training dataset, and the remaining 704 data as a testing dataset. The distribution of training data is illustrated in Figure 3.

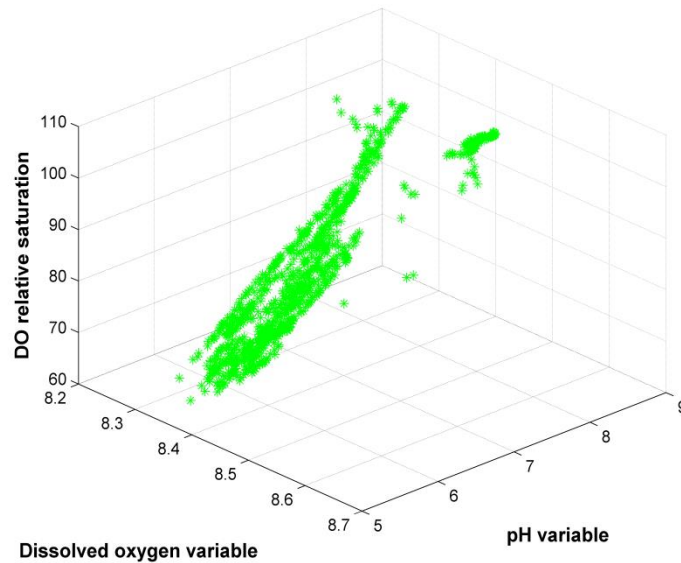


Figure 3. Illustration of the water quality dataset distribution in the training process. DO: dissolved oxygen.

Figure 3 is the training result distribution diagram of the ID-SVDD detection algorithm in the water quality dataset. In Figure 3, the green points represent all normal data in the training process. The three-dimensional coordinate represents the DO content, pH variable, and DO relative saturation variable, respectively. Most normal data are aggregated and distributed in an irregular shape. Small amounts of data are dispersedly distributed. The detection result on the testing dataset is shown in Figure 4.

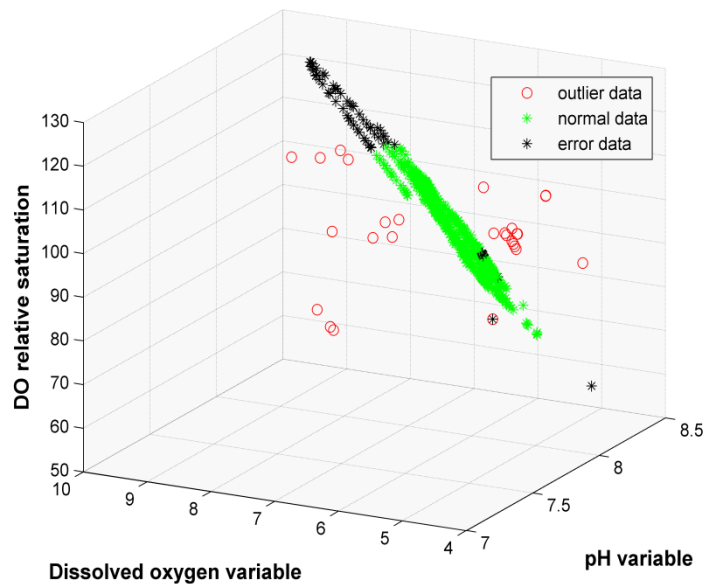


Figure 4. Detection results of the water quality dataset in the testing process.

We can see from Figure 4 that the three-dimensional coordinate is the same as Figure 3. After ID-SVDD outlier detection, the error points are shown in Figure 4 with the form of a black dot. Error points appear in both the normal dataset and the outlier dataset. The outlier data are distributed around the normal data. To evaluate the performance of the ID-SVDD algorithm, we made

a comparison with D-SVDD, DW-SVDD, and traditional SVDD. The precision comparison results are shown in Table 5. Figure 5 presents the run-time comparison of the four algorithms.

Table 5. Detection results for the water quality dataset. D-SVDD: density-compensated SVDD; DW-SVDD: density-weighted SVDD; ID-SVDD: improved density-compensated SVDD.

pond13	ID-SVDD	D-SVDD	DW-SVDD	SVDD
TPR (%)	91.1374	89.0694	70.901	67.356
TNR (%)	96.2963	100	92.5926	96.2963
Accuracy (%)	91.3352	89.4886	71.733	68.4659

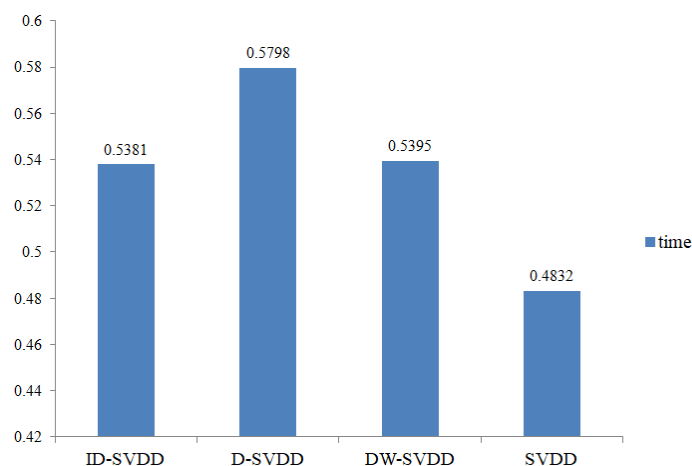


Figure 5. Outlier detection time of the water quality dataset with different algorithms.

It can be seen from Table 5 that ID-SVDD had the highest values of TPR and TNR, with 91.335% detection accuracy. The TPR of ID-SVDD was 2.322%, 28.542%, and 35.307% higher than those of D-SVDD, DW-SVDD, and traditional SVDD, respectively. The TNR of ID-SVDD was 4% greater than that of DW-SVDD, and equal to that of SVDD. ID-SVDD was successful in detecting the outliers of water quality data, increasing the accuracy by 2.064%, 27.327%, and 33.403% when compared to D-SVDD, DW-SVDD, and SVDD, respectively. There are correlations among pH, DO, and DO relative saturation. MD improved Parzen-window relative density weight can eliminate the interference of correlations, thus improving the detection performance. Meanwhile, the TPR, TNR, and accuracy of D-SVDD were superior to DW-SVDD and SVDD. That is because Parzen-window relative density weight can obtain a characterized description of the dataset in high-dimensional feature space and help search for an optimal SVDD. This approach is suitable for calculating the relative density weight. The introduction of improved relative density to SVDD helps enhance the performance for outlier detection efficiently.

As Figure 5 indicates, ID-SVDD had an advantage over D-SVDD and DW-SVDD in terms of run time. It consumed 0.5381 s for outlier detection. Single SVDD provided the shortest time (0.4832 s), but its TPR and accuracy were the lowest among the four algorithms. Therefore, ID-SVDD provided satisfactory outlier detection accuracy and efficiency, and it is suitable for detecting outliers in real water quality monitoring.

4. Conclusions

This paper presents a new outlier detection algorithm (ID-SVDD) incorporating the relative density weight with SVDD. This approach can obtain the features of the data, thus improving the performance of SVDD. To measure the relative density weight, we used the Parzen-window method. The Mahalanobis distance was applied to improve the Gaussian function in the calculation of relative density. ID-SVDD can realize data mapping from relatively sparse space to high-density space.

We conducted experiments to evaluate the performance of ID-SVDD based on SensorScope datasets and water quality datasets, and then compared it with D-SVDD, DW-SVDD, and SVDD algorithms. The experimental results showed that ID-SVDD performed better than its three other counterparts in terms of TPR, TNR, accuracy, and run time. Therefore, it is efficient and useful to introduce relative density to SVDD. ID-SVDD provides a new idea of outlier detection and it can be used in real-world applications.

Author Contributions: Conceptualization, L.K.; Data curation, Y.Y.; Investigation, Y.Y.; Methodology, P.S.; Software, L.K.; Validation, P.S.; Writing-original draft, P.S.; Writing-review and editing, G.L.

Funding: This research was funded in part by the National Natural Science Foundation of China (Grant No. 61472368), Central Public-interest Scientific Institution Basal Research Fund, CAFS (Grant No.2016HY-ZD1404), 111 Project (B12018), Key Research and Development Project of Jiangsu Province (Grant No. BE2016627), the Fundamental Research Funds for the Central Universities (Grant No. RP51635B), and Wuxi International Science and Technology Research and Development Cooperative Project (Grant No.CZE02H1706).

Acknowledgments: We thank the freshwater fisheries research center of Chinese Academy of Fishery Sciences for providing the aquaculture base.

Conflicts of Interest: The authors declare no conflicts of interest. The founding sponsors had no role in the design of the study, in the collection, analyses or interpretation of data, in the writing of the manuscript, nor in the decision to publish the results.

References

1. Gomes, R.D.; Queiroz, D.V.; Filho, A.C.L.; Fonseca, I.E.; Alencar, M.S. Real-time link quality estimation for industrial wireless sensor networks using dedicated nodes. *Ad Hoc Netw.* **2017**, *59*, 116–133. [[CrossRef](#)]
2. Periyanyagi, S.; Sumathy, V. Swarm-based defense technique for tampering and cheating attack in WSN using CPHS. *Pers. Ubiquitous Comput.* **2018**, *22*, 1165–1179. [[CrossRef](#)]
3. Alaiad, A.; Zhou, L. Patients' Adoption of WSN-Based Smart Home Healthcare Systems: An Integrated Model of Facilitators and Barriers. *IEEE Trans. Prof. Commun.* **2017**, *60*, 4–23. [[CrossRef](#)]
4. Khan, T.H.F.; Kumar, D.S. Ambient crop field monitoring for improving context based agricultural by mobile sink in WSN. *J. Ambient Intell. Humaniz. Comput.* **2019**, 1–9. [[CrossRef](#)]
5. Pierdicca, A.; Clementi, F.; Isidori, D.; Concettoni, E.; Cristalli, C.; Lenci, S. Numerical model upgrading of a historical masonry palace monitored with a wireless sensor network. *Int. J. Mason. Res. Innov.* **2016**, *1*, 74. [[CrossRef](#)]
6. Rainieri, C.; Fabbrocino, G. *Operational Modal Analysis of Civil Engineering Structures*; Springer: New York, NY, USA, 2014. [[CrossRef](#)]
7. Lynch, J.P. A Summary Review of Wireless Sensors and Sensor Networks for Structural Health Monitoring. *Shock Vib. Dig.* **2006**, *38*, 91–128. [[CrossRef](#)]
8. Federici, F.; Graziosi, F.; Faccio, M.; Colarieti, A.; Gattulli, V.; Lepidi, M.; Potenza, F. An Integrated Approach to the Design of Wireless Sensor Networks for Structural Health Monitoring. *Int. J. Distrib. Sens. Netw.* **2012**, *8*, 594842. [[CrossRef](#)]
9. Guang, X.Z.; Tian, W.; Guo, J.W.; An, F.L.; Wei, J.J. Detection of Hidden Data Attacks Combined Fog Computing and Trust Evaluation Method in Sensor-Cloud System. *Concurr. Comput. Pract. Exp.* **2018**. [[CrossRef](#)]
10. You, K.W.; Hai, Y.H.; Qun, W.; An, F.L.; Tian, W. A Risk Defense Method Based on Microscopic State Prediction with Partial Information Observations in Social Networks. *J. Parallel Distrib. Comput.* **2019**, *131*, 189–199.
11. Wang, T.; Zhou, J.; Liu, A.; Bhuiyan, M.Z.A.; Wang, G.; Jia, W. Fog-based Computing and Storage Offloading for Data Synchronization in IoT. *IEEE Internet Things J.* **2019**, *6*, 4272–4282. [[CrossRef](#)]
12. Wang, T.; Zhang, G.; Liu, A.; Bhuiyan, M.Z.A.; Jin, Q. A Secure IoT Service Architecture with an Efficient Balance Dynamics Based on Cloud and Edge Computing. *IEEE Internet Things J.* **2019**, *6*, 4831–4843. [[CrossRef](#)]
13. Ghosal, A.; Halder, S.; Dasbit, S. A dynamic TDMA based scheme for securing query processing in WSN. *Wirel. Netw.* **2012**, *18*, 165–184. [[CrossRef](#)]

14. Knorr, E.M.; Ng, R.T.; Tucakov, V. Distance-based outliers: Algorithms and applications. *VLDB J.* **2000**, *8*, 237–253. [[CrossRef](#)]
15. Sheng, B.; Li, Q.; Mao, W.; Jin, W. Outlier detection in sensor networks. In Proceedings of the 8th ACM International Symposium on Mobile and Ad Hoc Networking and Computing (MobiHoc), Montreal, QC, Canada, 9–14 September 2007; pp. 219–228.
16. Chen, Y.; Miao, D.; Zhang, H. Neighborhood outlier detection. *Expert Syst. Appl.* **2010**, *37*, 8745–8749. [[CrossRef](#)]
17. Xie, M.; Hu, J.; Han, S.; Chen, H.H. Scalable hypergrid k-nn-based online anomaly detection in wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 1661–1670. [[CrossRef](#)]
18. Shamshirband, S.; Amini, A.; Anuar, N.B.; Mat Kiah, M.L.; Teh, Y.W.; Furnell, S. D-FICCA: A density-based fuzzy imperialist competitive clustering algorithm for intrusion detection in wireless sensor networks. *Measurement* **2014**, *55*, 212–226. [[CrossRef](#)]
19. Wazid, M.; Das, A.K. An Efficient Hybrid Anomaly Detection Scheme Using K-Means Clustering for Wireless Sensor Networks. *Wirel. Pers. Commun.* **2016**, *90*, 1971–2000. [[CrossRef](#)]
20. Rajasegarar, S.; Leckie, C.; Palaniswami, M. Hyperspherical cluster based distributed anomaly detection in wireless sensor networks. *J. Parallel Distrib. Comput.* **2014**, *74*, 1833–1847. [[CrossRef](#)]
21. Moshtaghi, M.; Leckie, C.; Karunasekera, S.; Rajasegarar, S. An adaptive elliptical anomaly detection model for wireless sensor networks. *Comput. Netw.* **2014**, *64*, 195–207. [[CrossRef](#)]
22. Hill, D.J.; Minsker, B.S.; Amir, E. Real-time bayesian anomaly detection in streaming environmental data. *Water Resour. Res.* **2009**, *45*, 450–455. [[CrossRef](#)]
23. Kang, L.; Xu, L.; Zhao, J. Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model. *IEEE Trans. Knowl. Data Eng.* **2018**, *27*, 636–650.
24. Yue, R.; Xue, X.; Liu, H.; Tan, J.; Li, X. Quantum Algorithm for K-Nearest Neighbors Classification Based on the Metric of Hamming Distance. *Int. J. Theor. Phys.* **2017**, *56*, 1–12.
25. Tao, J.; Dan, H.; Yu, X. Enhanced IT2FCM algorithm using object-based triangular fuzzy set modeling for remote-sensing clustering. *Comput. Geosci.* **2018**, *118*, 14–26.
26. Tax, D.; Duin, R. Support vector domain description. *Pattern Recognit. Lett.* **1999**, *20*, 1191–1199. [[CrossRef](#)]
27. Shin, J.H.; Lee, B.; Park, K.S. Detection of abnormal living patterns for elderly living alone using support vector data description. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 438–448. [[CrossRef](#)]
28. Bovolo, F.; Camps-Valls, G.; Bruzzone, L. A support vector domain method for change detection in multitemporal images. *Pattern Recognit. Lett.* **2010**, *31*, 1148–1154. [[CrossRef](#)]
29. Khediri, I.B.; Weihs, C.; Limam, M. Kernel k-means clustering based local support vector domain description fault detection of multimodal processes. *Expert Syst. Appl.* **2012**, *39*, 2166–2171. [[CrossRef](#)]
30. Liu, Y.H.; Liu, Y.C.; Chen, Y.Z. High-speed inline defect detection for TFT-LCD array process using a novel support vector data description. *Expert Syst. Appl.* **2011**, *38*, 6222–6231. [[CrossRef](#)]
31. Zhao, Y.; Wang, S.; Xiao, F. Pattern recognition-based chillers fault detection method using support vector data description (SVDD). *Appl. Energy* **2013**, *112*, 1041–1048. [[CrossRef](#)]
32. Lee, K.; Kim, D.W.; Lee, K.H.; Lee, D. Density-induced support vector data description. *IEEE Trans. Neural Netw.* **2007**, *18*, 284–289. [[CrossRef](#)]
33. Cha, M.; Kim, J.S.; Baek, J.G. Density weighted support vector data description. *Expert Syst. Appl.* **2014**, *41*, 3343–3350. [[CrossRef](#)]
34. Sun, R.; Tsung, F.A. Kernel-distance-based multivariate control charts using support vector methods. *Int. J. Prod. Res.* **2003**, *41*, 2975–2989. [[CrossRef](#)]
35. Zhou, Y.; Wu, K.; Meng, Z.; Tian, M. Fault detection of aircraft based on support vector domain description. *Comput. Electr. Eng.* **2017**, *61*, 80–94. [[CrossRef](#)]
36. Belghith, A.; Bowd, C.; Medeiros, F.A.; Balasubramanian, M.; Weinreb, R.N.; Zangwill, L.M. Learning from healthy and stable eyes: A new approach for detection of glaucomatous progression. *Artif. Intell. Med.* **2015**, *64*, 105–115. [[CrossRef](#)]
37. Wu, Y.; Chen, P.; Yao, Y.; Ye, X.; Xiao, Y.; Liao, L.; Wu, M.; Chen, J. Dysphonic voice pattern analysis of patients in parkinson’s disease using minimum interclass probability risk feature selection and bagging ensemble learning methods. *Comput. Math. Methods Med.* **2017**, *2017*, 4201984. [[CrossRef](#)]
38. Bothorel, C.; Cruz, J.D.; Magnani, M.; Mícenková, B. Clustering attributed graphs: Models, measures and methods. *Netw. Sci.* **2015**, *3*, 408–444. [[CrossRef](#)]

39. Long, B.; Tian, S.; Wang, H. Feature vector selection method using mahalanobis distance for diagnostics of analog circuits based on ls-svm. *J. Electron. Test.* **2012**, *28*, 745–755. [[CrossRef](#)]
40. Jiang, C.; Zhang, S.B. A Novel Adaptively-Robust Strategy Based on the Mahalanobis Distance for GPS/INS Integrated Navigation Systems. *Sensors* **2018**, *18*, 695. [[CrossRef](#)]
41. Sensor Scope Sytem [DB/OL]. Available online: <http://sensorscope.epfl.ch/index.php/MainPage> (accessed on 19 December 2016).
42. Luo, Y.; Li, Z.; Guo, H.; Cao, H.; Song, C.; Guo, X.; Zhang, Y. Predicting congenital heart defects: A comparison of three data mining methods. *PLoS ONE* **2017**, *12*, e0177811. [[CrossRef](#)]
43. Tian, W.; Hao, L.; James, X.Z.; Mande, X. Crowdsourcing Mechanism for Trust Evaluation in CPCS based on Intelligent Mobile Edge Computing. *ACM Trans. Intell. Syst. Technol.* **2019**. [[CrossRef](#)]
44. Zhang, Y.; Guo, H.; Chen, F.; Yang, H. Weighted kernel mapping model with spring simulation based watershed transformation for level set image segmentation. *Neurocomputing* **2017**, *249*, 1–18. [[CrossRef](#)]
45. Duan, L.; Zhang, H.; Khan, M.S.; Fang, M. A self-adaptive frequency selection common spatial pattern and least squares twin support vector machine for motor imagery electroencephalography recognition. *Biomed. Signal Process. Control* **2018**, *41*, 222–232.
46. Shi, P.; Li, G.; Yuan, Y.; Huang, G.; Kuang, L. Prediction of dissolved oxygen content in aquaculture using clustering-based softplus extreme learning machine. *Comput. Electron. Agric.* **2019**, *157*, 329–338. [[CrossRef](#)]



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