

Research Article

An Improved Semisupervised Outlier Detection Algorithm Based on Adaptive Feature Weighted Clustering

Tingquan Deng^{1,2} and Jinhong Yang²

¹College of Science, Harbin Engineering University, Harbin 150001, China

²College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

Correspondence should be addressed to Jinhong Yang; yangjinhong.66@163.com

Received 13 September 2016; Revised 2 November 2016; Accepted 20 November 2016

Academic Editor: Filippo Ubertini

Copyright © 2016 T. Deng and J. Yang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There exist already various approaches to outlier detection, in which semisupervised methods achieve encouraging superiority due to the introduction of prior knowledge. In this paper, an adaptive feature weighted clustering-based semisupervised outlier detection strategy is proposed. This method maximizes the membership degree of a labeled normal object to the cluster it belongs to and minimizes the membership degrees of a labeled outlier to all clusters. In consideration of distinct significance of features or components in a dataset in determining an object being an inlier or outlier, each feature is adaptively assigned different weights according to the deviation degrees between this feature of all objects and that of a certain cluster prototype. A series of experiments on a synthetic dataset and several real-world datasets are implemented to verify the effectiveness and efficiency of the proposal.

1. Introduction

Outlier detection is an important topic in data mining community, which aims at finding patterns that occur infrequently as opposed to other data mining techniques [1]. An outlier is an observation that deviates significantly from, or inconsistent with the main body of a dataset, as if it was generated by a different mechanism [2]. The importance of outlier detection is in the view of the fact that outliers can provide raw patterns and valuable knowledge about a dataset. Current application areas of outlier detection include crime detection, credit card fraud detection, network intrusion detection, medical diagnosis, faulty detection in critical safety systems, or detecting abnormal regions in image processing [3–9].

Recently the studies on outlier detection are very active and many approaches have been proposed. In general, existing work on outlier detection can be broadly classified into three modes depending on whether label information is available or can be used to build outlier detection models: unsupervised, supervised, and semisupervised methods.

Supervised outlier detection concerns the situation where the training dataset contains prior information about the class of each instance that is normal or abnormal. One-class support vector machine (OCSVM) [10] or support

vector data description (SVDD) [11, 12], which considers the case that training data are all normal instances, conducts a hypersphere around the normal data and utilizes the constructed hypersphere to detect an unknown sample as an inlier or outlier. The supervised outlier detection problem is a difficult case in many real-world applications, since the acquisition of label information of the whole training dataset is often expensive, time consuming, and subjective.

Unsupervised outlier detection, without prior information about the class distribution, is generally classified into distribution-based [3], distance-based [13, 14], density-based [15, 16], and clustering-based [17–20] approaches. Distribution-based approach assumes that all data points are generated by a certain statistical model, while outliers do not obey the model. However, the assumption of an underlying distribution of data points is not always available in many real-life applications. Distance-based approach was firstly investigated by Knox and Ng [14]. An object o in a dataset X is an outlier if at least $p\%$ of objects in X are further than the distance D from o . The global parameters p and D are not suitable when the local information of the dataset varies greatly. Representatives of this type of approaches include K -nearest neighbor (KNN) algorithm [13] and its variants [21, 22]. Density-based approach was originally proposed by

Breunig et al. [15]. A local outlier factor (LOF) is assigned to each data point based on their local neighborhood density. Then a data point with a high LOF value is determined as an outlier. However, this method is very sensitive to the choice of neighborhood parameter.

Clustering-based approaches [17–20] partition the dataset into several clusters depending on similarity of objects and detect outliers by examining the relationship between objects and clusters. In general, clusters containing significantly less data points than other clusters or being remote from other clusters are considered as outliers. The cluster structure of data can facilitate the task of outlier detection and a small amount of related literatures has been proposed. A classical clustering method is used to find anomaly in the intrusion detection domain [18]. In the work of [19], the clustering techniques iteratively detect outliers for multidimensional data analysis in subspace. Zhao et al. [20] propose an adaptive fuzzy *c*-means (AFCM) algorithm by introducing sample weight coefficients to the objective function and apply it to anomaly data detection in energy system of steel industry. Since clustering-based approaches are unsupervised without requiring any labeled training data, their performance in outlier detection is limited. In addition, most of the existing clustering-based methods only involve the optimal clustering but do not incorporate optimal outlier detection into clustering process.

In many real-world applications, one may encounter cases where a small set of objects are labeled as outliers or belonging to a certain class, but most of the data are unlabeled. Studies indicate that the introduction of a small amount of prior knowledge can significantly improve the effectiveness of outlier detection [23–25]. Therefore, semisupervised approaches to outlier detection have been developed to tackle such scenarios and have been thought of a popular direction of outlier detection recently. In order to take advantage of the label information of a target dataset, entropy-based outlier detection based on semisupervised learning from few positive examples (EODSP) is proposed in [23]. That method extracts reliable normal instances from unlabeled objects and regards them as labeled normal samples. Entropy-based outlier detection method is used to detect top M outliers. However, when a dataset initially provides labeled normal and abnormal samples, the algorithm in [23] cannot make full use of the given label information. Literature [24] develops a semisupervised outlier detection method based on the assessment of deviation from known labeled objects by punishing poor clustering results and restricting the number of outliers. Xue et al. [25] present a semisupervised outlier detection proposal based on fuzzy rough *c*-means clustering, which detects outliers by minimizing the sum of squared errors of clustering results and the deviation from known labeled examples as well as the number of outliers. Unfortunately, some labeled normal objects are finally misidentified as outliers due to improper parameter selection in [24, 25].

Most of the previous research equally treats different features of objects in outlier detecting process, which does not conform to the intrinsic characteristic of a dataset. Actually, it is more reasonable that different features have different importance in each cluster, especially for high-dimension

sparse datasets where the structure of each cluster is often limited to a subset of features rather than the entire feature set. Some works concerning feature weighted clustering have been studied. Huang et al. [26] propose a *W*-*c*-means type clustering algorithm that can automatically calculate feature weights. *W*-*c*-means adds a new step into the basic *c*-means algorithm to update the variable weights based on the current partition of data. Literature [27] develops an approach called simultaneous clustering and attribute discrimination (SCAD). SCAD learns the feature relevance representation of each cluster independently in an unsupervised manner. Zhou et al. [28] publish a maximum-entropy-regularized weighted fuzzy *c*-means (EWFCM) clustering algorithm for “non-spherical” shaped data. A new objective function is developed in the EWFCM algorithm to achieve the optimal clustering result by minimizing the dispersion within clusters and maximizing the entropy of attribute weights simultaneously. These existing methods about feature weighted clustering encourage scholars to study outlier detection based on feature weighted clustering.

To make full use of prior knowledge to facilitate clustering-based outlier detection, we develop a semisupervised outlier detection algorithm based on adaptive feature weighted clustering (SSOD-AFW) in this paper, in which the feature weights are iteratively obtained. The proposed algorithm emphasizes the diversity of different features in each cluster and assigns lower weights to irrelevant features to reduce their negative influence on outlier decision. Furthermore, based on the convention that outliers usually have a lower membership to every cluster, we relax the constraint of fuzzy *c*-means (FCM) clustering where the membership degrees of a sample to all clusters must sum up to one and propose an adaptive feature weighted semisupervised possibilistic clustering-based outlier detection algorithm. The interaction problem between optimal clustering and outlier detection is addressed in the proposed method. The label information is introduced into the possibilistic clustering method according to the following principles: (1) maximizing the membership degree of a labeled normal object to the cluster it belongs to; (2) minimizing the membership degrees of a labeled normal object to the clusters it does not belong to; and (3) minimizing the membership degrees of a labeled outlier to all clusters. In addition to the above principles, we simultaneously minimize the dispersion within clusters in the new objective function of clustering to achieve a proper cluster structure. Finally the yielded optimal membership degrees are used to indicate the outlying degree of each sample in the dataset. The proposed algorithm is found promising in improving the performance of outlier detection in comparison with typical outlier detection methods in accuracy, running time as well as other evaluation metrics.

The remainder of this paper is organized as follows. Section 2 gives a short review on possibilistic clustering algorithms. Section 3 presents the detailed description of feature weighted semisupervised clustering-based outlier detection algorithm. In Section 4, the experimental results of the proposed method against typical outlier detection algorithms are discussed on synthetic and real-world datasets. Finally, Section 5 follows our conclusions.

2. Possibilistic Clustering Algorithms

Let $X = \{x_1, x_2, \dots, x_n\}$ be a given dataset of n objects, where $x_i = \{x_{i1}, x_{i2}, \dots, x_{iN}\} \in R^N$ is the i th object characterized by N features. Suppose that the dataset is divided into c clusters and o_k denotes the k th cluster prototype.

FCM is a well-known clustering algorithm [29], whose objective function is

$$J_{\text{FCM}} = \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m \|x_i - o_k\|^2, \quad (1)$$

$$\text{s.t. } \sum_{k=1}^c u_{ik} = 1, \quad u_{ik} \in [0, 1], \quad (2)$$

where u_{ik} is the membership degree of the i th ($1 \leq i \leq n$) object to the k th ($1 \leq k \leq c$) cluster. $\|\cdot\|$ represents the l_2 -norm of a vector and $m > 0$ is the fuzzification coefficient. Note that the constraint condition in (2) indicates that the membership sum of each object to all clusters equals one. Therefore, FCM is sensitive to outliers due to the intuition that outliers or noises commonly locate far away from all cluster prototypes. For this reason, Krishnapuram and Keller [30] proposed a possibilistic c -means (PCM) clustering algorithm, which relaxes the constraint on the sum of memberships and minimizes the following objective function:

$$J_{\text{PCM}} = \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m \|x_i - o_k\|^2 + \sum_{k=1}^c \eta_k \sum_{i=1}^n (1 - u_{ik})^m, \quad (3)$$

$$\text{s.t. } u_{ik} \in [0, 1], \quad (4)$$

where η_k is a suitable positive number. In PCM, the constraint (4) allows an outlier holding a low membership to all clusters, so an outlier has a low impact on the objective function (3). The membership information of each sample can be naturally used to interpret the outlying characteristic of a sample. For a certain sample, if it has a low membership to all clusters, it is likely to be an outlier.

Afterward, another unsupervised possibilistic clustering algorithm (PCA) is proposed by Yang and Wu [31] and the objective function of PCA is described as

$$\begin{aligned} J_{\text{PCA}} &= \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m \|x_i - o_k\|^2 \\ &+ \frac{\beta}{m^2 \sqrt{c}} \sum_{i=1}^n \sum_{k=1}^c (u_{ik}^m \log u_{ik}^m - u_{ik}^m), \end{aligned} \quad (5)$$

$$\text{s.t. } u_{ik} \in [0, 1],$$

where the parameter β can be calculated by the sample covariance:

$$\beta = \frac{\sum_{i=1}^n \sum_{j=1}^N (x_{ij} - \sum_{l=1}^n x_{lj}/n)^2}{n}. \quad (6)$$

3. Semisupervised Outlier Detection Framework Based on Feature Weighted Clustering

3.1. Model Formulation. In this section, we introduce prior knowledge into possibilistic c -means clustering method to improve the performance of outlier detection. First, a small subset of samples in a given dataset $X = \{x_1, x_2, \dots, x_n\}$ is labeled as normal or outlier objects. Each labeled normal object carries the label of class it belongs to. A semisupervised indicator matrix $\mathbf{A} = (a_{ik})_{n \times c}$ is constructed to describe the semisupervised information and its entries are defined by the following:

- (i) If an object x_i is labeled as a normal point and it belongs to the k th cluster, then $a_{ik} = -1$, and for all $s = 1, 2, \dots, c$, $s \neq k$, we let $a_{is} = 1$.
- (ii) If x_i is labeled as an outlier, then for all $s = 1, 2, \dots, c$, we set $a_{is} = 1$.
- (iii) If x_i is unlabeled, then for all $s = 1, 2, \dots, c$, it has $a_{is} = 0$.

Usually data often contain a number of redundant features. The cluster structure in a given dataset is often confined to a subset of features rather than the entire feature set. Irrelevant features can only obscure the discovery of the cluster structure by a clustering algorithm. An intrinsic outlier is easy to be neglected due to the vagueness of cluster structure. Figure 1 presents an example of a three-dimensional dataset. The dataset has two clusters (c_1 and c_2) and 3 features (f_1, f_2 , and f_3). In the feature space (f_1, f_2, f_3), neither of the clusters is discovered (see Figure 1(a)). In the subspace (f_1, f_2), cluster c_1 can be found, but c_2 cannot (see Figure 1(b)). Nevertheless, only cluster c_2 can be clearly shown in (f_2, f_3) (see Figure 1(c)). Therefore, if we assign weights 0.47, 0.45, and 0.08 to features f_1, f_2 , and f_3 , respectively, cluster c_1 will be recovered by a clustering algorithm. If the weights of features f_1, f_2 , and f_3 are assigned as 0.13, 0.46, and 0.41, respectively, cluster c_2 will be recovered. In this consideration, each cluster is relevant to different subsets of features, and the same feature may have different importance in different clusters.

In our research, let v_{jk} be the weight of the j th ($1 \leq j \leq N$) dimensional feature with respect to the k th ($1 \leq k \leq c$) cluster, which satisfies $\sum_{j=1}^N v_{jk} = 1$; then the feature weighted distance d_{ik} between the i th object and the k th cluster prototype is defined as

$$d_{ik} = \sqrt{\sum_{j=1}^N v_{jk}^q (x_{ij} - o_{kj})^2}, \quad (7)$$

where the parameter $q > 1$ is the feature weight index.

The points within clusters usually behave strongly correlated, while weak correlation is shown between outliers. That is, normal points belong to one of the c clusters and outliers do not belong to any cluster. Therefore, a normal point should have a high membership to the cluster it belongs to, and an outlier has a low membership to all clusters. Based on this

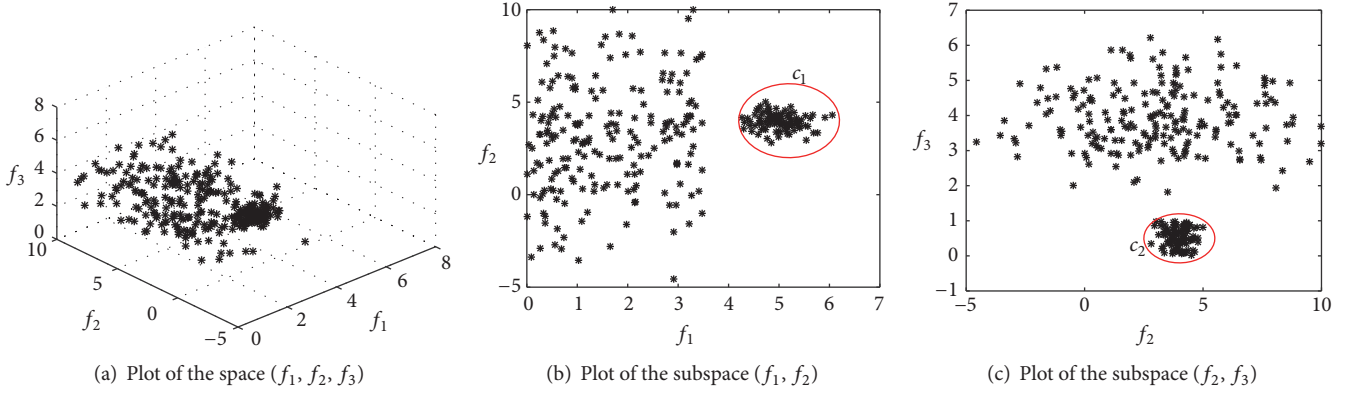


FIGURE 1: Three-dimensional synthetic example.

idea, we define a new objective function and minimize it as follows:

$$\begin{aligned}
 J_{\text{SSOD-AFW}}(\mathbf{U}, \mathbf{O}, \mathbf{V}) &= \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m d_{ik}^2 + \frac{\beta}{m^2 \sqrt{c}} \sum_{i=1}^n \sum_{k=1}^c (u_{ik}^m \log u_{ik}^m - u_{ik}^m) \\
 &\quad + \sum_{i=1}^n \sum_{k=1}^c \alpha_{ik} a_{ik} u_{ik}^m, \quad (8)
 \end{aligned}$$

$$\text{s.t. } \sum_{j=1}^N v_{jk} = 1, \quad u_{ik} \in [0, 1], \quad v_{jk} \in [0, 1], \quad (9)$$

where n , N , and c are the number of objects, features, and clusters, respectively. $\mathbf{U} = (u_{ik})_{n \times c}$, u_{ik} is the membership degree of the i th object belonging to the k th cluster. $\mathbf{V} = (v_{jk})_{N \times c}$, v_{jk} denotes the feature weight of the j th dimensional feature with respect to the k th cluster. $\mathbf{O} = (o_{kj})_{c \times N}$, o_{kj} indicates the j th dimensional feature value of the k th cluster prototype. d_{ik} denotes the feature weighted distance between object x_i and the k th cluster prototype. $a_{ik} \in \{1, -1, 0\}$ is the element in semisupervised indicator matrix \mathbf{A} . $m > 0$ is the fuzzification coefficient and the parameter β can be fixed as the sample covariance according to (6). The positive coefficient α_{ik} adjusts the significance of the label information of the i th object with respect to the k th cluster in objective function (8). The larger α_{ik} is, the larger the influence of label knowledge is.

The first term in (8) is equivalent to the FCM objective function which requires the distances of objects from the cluster prototypes to be as small as possible. The second term is constructed to force u_{ik} to be as large as possible. The third term $\alpha_{ik} a_{ik} u_{ik}^m$ focuses on minimizing the membership degrees of a labeled outlier to all the clusters and maximizing the membership degree of a labeled normal object to the cluster it belongs to. With a proper choice of α_{ik} , we can balance the label information weight of every object and achieve the optimal fuzzy partition.

The virtue of semisupervised indicator matrix \mathbf{A} in objective function (8) can be elaborated as follows. Recalling the construction of semisupervised indicator matrix \mathbf{A} and

objective function (8), note that if we know that x_i belongs to the k th cluster, then $a_{ik} = -1$ and all the other entries in the i th row equal 1. Thus, minimizing $\alpha_{ik} a_{ik} u_{ik}^m$ in (8) means maximizing the membership of x_i to the k th cluster and simultaneously minimizing the memberships of x_i to the other clusters. If x_i is labeled as an outlier, namely, where all the elements in the i th row of \mathbf{A} equal 1, then minimizing $\alpha_{ik} a_{ik} u_{ik}^m$ in (8) means minimizing the memberships of x_i to all clusters, for an outlier does not belong to any cluster. If x_i is unlabeled, namely, where $a_{ik} = 0$ for all $k = 1, 2, \dots, c$, then the term $\alpha_{ik} a_{ik} u_{ik}^m$ has no impact on objective function (8).

3.2. Solutions to the Objective Function. In this subsection, an iterative algorithm for minimizing $J_{\text{SSOD-AFW}}$ with respect to u_{ik} , v_{jk} , and o_{kj} is derived similar to classical FCM.

First, in order to minimize $J_{\text{SSOD-AFW}}$ with respect to \mathbf{V} , \mathbf{U} and \mathbf{O} are fixed and the parameters α_{ik} ($i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$) are constants. The Lagrange function is constructed as follows:

$$\begin{aligned}
 L_1 &= \sum_{i=1}^n \sum_{k=1}^c u_{ik}^m d_{ik}^2 + \frac{\beta}{m^2 \sqrt{c}} \sum_{i=1}^n \sum_{k=1}^c (u_{ik}^m \log u_{ik}^m - u_{ik}^m) \\
 &\quad + \sum_{i=1}^n \sum_{k=1}^c \alpha_{ik} a_{ik} u_{ik}^m - \sum_{k=1}^c \gamma_k \left(\sum_{j=1}^N v_{jk} - 1 \right), \quad (10)
 \end{aligned}$$

where γ_k ($k = 1, 2, \dots, c$) are the Lagrange multipliers.

By taking the gradient of L_1 with respect to v_{jk} and setting it to zero, we obtain

$$\frac{\partial L_1}{\partial v_{jk}} = q v_{jk}^{q-1} \sum_{i=1}^n u_{ik}^m (x_{ij} - o_{kj})^2 - \gamma_k = 0. \quad (11)$$

Then

$$v_{jk} = \left[\frac{\gamma_k}{q \sum_{i=1}^n u_{ik}^m (x_{ij} - o_{kj})^2} \right]^{1/(q-1)}. \quad (12)$$

Substituting (12) into (9), we have

$$\sum_{j=1}^N v_{jk} = \left[\frac{\gamma_k}{q} \right]^{1/(q-1)} \sum_{j=1}^N \left[\frac{1}{\sum_{i=1}^n u_{ik}^m (x_{ij} - o_{kj})^2} \right]^{1/(q-1)} \quad (13)$$

$$= 1.$$

It follows that

$$\left[\frac{\gamma_k}{q} \right]^{1/(q-1)} = \frac{1}{\sum_{j=1}^N \left[1/\sum_{i=1}^n u_{ik}^m (x_{ij} - o_{kj})^2 \right]^{1/(q-1)}}. \quad (14)$$

The updating criteria of feature weight v_{jk} ($1 \leq j \leq N$, $1 \leq k \leq c$) are obtained:

$$v_{jk} = \frac{\left[1/\sum_{i=1}^n u_{ik}^m (x_{ij} - o_{kj})^2 \right]^{1/(q-1)}}{\sum_{l=1}^N \left[1/\sum_{i=1}^n u_{il}^m (x_{il} - o_{kl})^2 \right]^{1/(q-1)}}. \quad (15)$$

The updating way of v_{jk} implies that the larger the deviation degrees from all samples to the k th cluster prototype regarding the j th feature are, the smaller the weight of the j th feature is. That is, if the distribution of all data is compact around the k th cluster prototype in the j th feature space, the j th feature plays a significant role in formulating the k th cluster. Meanwhile, irrelevant features thus are assigned a smaller weight to reduce the negative impact of them on the clustering process.

To find the optimal cluster prototype \mathbf{O} , we assume \mathbf{U} and \mathbf{V} are fixed and the parameters α_{ik} ($i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$) are also constants. We take the gradient of $J_{\text{SSOD-AFW}}$ with respect to o_{kj} and set it to zero:

$$\frac{\partial J_{\text{SSOD-AFW}}}{\partial o_{kj}} = \sum_{i=1}^n (-2u_{ik}^m v_{jk} (x_{ij} - o_{kj})) = 0. \quad (16)$$

The updating formula of cluster prototype o_{kj} is obtained as follows:

$$o_{kj} = \frac{\sum_{i=1}^n u_{ik}^m x_{ij}}{\sum_{i=1}^n u_{ik}^m}. \quad (17)$$

To solve the optimal fuzzy partition matrix \mathbf{U} , we assume \mathbf{O} and \mathbf{V} are fixed and the parameters α_{ik} ($i = 1, 2, \dots, n$; $k = 1, 2, \dots, c$) are also constants. We set the gradient of $J_{\text{SSOD-AFW}}$ with respect to u_{ik} to zero:

$$\frac{\partial J_{\text{SSOD-AFW}}}{\partial u_{ik}} = m u_{ik}^{m-1} d_{ik}^2 + \frac{\beta}{m \sqrt{c}} u_{ik}^{m-1} \log u_{ik}^m + \alpha_{ik} a_{ik} m u_{ik}^{m-1} = 0. \quad (18)$$

The updating formula of u_{ik} is derived as follows:

$$u_{ik} = \exp \left(\frac{-m \sqrt{c}}{\beta} (d_{ik}^2 + \alpha_{ik} a_{ik}) \right). \quad (19)$$

Formula (19) indicates that a large value of weighted distance d_{ik} leads to a smaller value of u_{ik} , for all $1 \leq i \leq n$,

$1 \leq k \leq c$. It should be noted that the membership degree u_{ik} is also dependent on the coefficient α_{ik} . The choice of α_{ik} is important to the performance of the SSOD-AFW algorithm because it serves in distinguishing the importance of the third term relative to the other terms in objective function (8). If α_{ik} is too small, the third term will be neglected and the labels of objects will not work to promote the cluster structure. If α_{ik} is too large, the other terms will be neglected, and the negative influence of possible mislabels of objects will be enlarged. The value of α_{ik} should be chosen such that it has the same order of magnitude with the first term in (8). To determine the parameter α_{ik} in an adaptive way, in all experiments described in this paper, we choose α_{ik} proportional to d_{ik}^2 as follows:

$$\alpha_{ik} = K d_{ik}^2, \quad (20)$$

where $K \in (0, 1)$ is a constant. Since the weighted distance d_{ik} is dynamically updated, the value of parameter α_{ik} is adaptively updated in each iteration.

3.3. Criteria for Outlier Identification. Based on the above analysis, outliers should hold low membership degrees to all clusters. Therefore, the sum of memberships of an object to all clusters can be used to evaluate its outlying degree. For a certain object x_i , its outlying degree is defined as

$$\text{OD}(x_i) = \sum_{k=1}^c u_{ik}. \quad (21)$$

Thus, a small value of $\text{OD}(x_i)$ indicates a high outlying possibility of object x_i . The outlying degree of each sample in a dataset is calculated, respectively, and sorted incrementally. The suspicious outliers can be found just by extracting the top M objects in the sorted outlying degree sequence, where M is a given number of outliers contained in the dataset or a given number of outliers one needs.

In summary, the description of the SSOD-AFW algorithm is shown in Algorithm 1.

Algorithm 1 (semisupervised outlier detection based on adaptive feature weighted clustering (SSOD-AFW)).

Input. Dataset X , the label information of some objects, the number of clusters c , parameters $m > 1$, $K \in (0, 1)$, $q > 1$, $\varepsilon > 0$, and the number of outliers M .

Output. M suspicious outliers.

- (1) Calculate the parameter β according to (6), randomly initialize matrix $\mathbf{U}^{(0)} = (u_{ik}^{(0)})_{n \times c}$, and initialize all elements in $\mathbf{V}^{(0)} = (v_{jk}^{(0)})_{N \times c}$ as $1/N$. Set iteration counter $t = 0$.
- (2) Compute the matrix of cluster prototype $\mathbf{O}^{(t+1)} = (o_{kj}^{(t+1)})_{c \times N}$ according to (17).
- (3) Update the feature weight matrix $\mathbf{V}^{(t+1)} = (v_{jk}^{(t+1)})_{N \times c}$ by (15).
- (4) Update the feature weighted distance $\mathbf{D}^{(t+1)} = (d_{ik}^{(t+1)})_{n \times c}$ by (7).

- (5) Update parameter $\alpha_{ik}^{(t+1)}$ by (20).
- (6) Update the membership degree matrix $\mathbf{U}^{(t+1)} = (u_{ik}^{(t+1)})_{n \times c}$ according to (19).
- (7) If $\|\mathbf{U}^{(t+1)} - \mathbf{U}^{(t)}\| < \varepsilon$, go to step (8); else, $t = t + 1$, repeat step (2) to step (7).
- (8) The outlying degree of each object is computed, and the OD values are sorted in an ascending manner. Finally output top M outliers with the smallest outlying degrees.

Computational complexity analysis: Step (2) needs $O(cNn)$ operations to compute c cluster prototype. The computational complexity of computing the weights of N features is $O(cNn)$ in Step (3). Step (4) requires $O(cNn)$ to compute the weighted distances of n objects to c cluster prototypes. Step (5) needs $O(cn)$ to compute parameter α_{ik} of n objects with respect to c cluster prototypes. Moreover, Step (6) needs $O(cn)$ operations to calculate the memberships of n objects to c clusters. Therefore, the whole computational complexity is $O(cNn)$, the same as that of the classical FCM algorithm.

3.4. Proof of Convergence. In this section, we discuss the convergence of the proposed SSOD-AFW algorithm. To prove the convergence of objective function $J_{\text{SSOD-AFW}}$ in (8) by iterating \mathbf{V} , \mathbf{O} , and \mathbf{U} with formulas (15), (17), and (19), it only needs to prove that $J_{\text{SSOD-AFW}}$ is monotonically decreasing and bounded after a finite number of iterations. Next Lemmas 2, 3, and 4 verify the monotonically decreasing property of $J_{\text{SSOD-AFW}}$ with respect to \mathbf{V} , \mathbf{O} , and \mathbf{U} , respectively. Lemma 5 presents the boundedness of $J_{\text{SSOD-AFW}}$.

Lemma 2. *Objective function $J_{\text{SSOD-AFW}}$ in (8) is nonincreasing by updating $\mathbf{U} = (u_{ik})_{n \times c}$ with formula (19).*

Proof. Due to the fact that \mathbf{V} and \mathbf{O} are fixed when updating \mathbf{U} by (19), here the objective function $J_{\text{SSOD-AFW}}$ can be regarded as a function only associated with \mathbf{U} , denoted as $J_{\text{SSOD-AFW}}(\mathbf{U})$. According to Lagrangian multiplier technique, \mathbf{U}^* computed via (19) is a stagnation point of $J_{\text{SSOD-AFW}}(\mathbf{U})$. On the other hand, if Hessian matrix $\nabla^2 J_{\text{SSOD-AFW}}(\mathbf{U})$ is proved to be positive definite at \mathbf{U}^* , it can be proved that $J_{\text{SSOD-AFW}}(\mathbf{U})$ attains its local minimum at \mathbf{U}^* . The Hessian matrix $\nabla^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*)$ is expressed as

$$\nabla^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*) = \begin{bmatrix} \frac{\partial^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*)}{\partial u_{11} \partial u_{11}} & \cdots & \frac{\partial^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*)}{\partial u_{11} \partial u_{nc}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*)}{\partial u_{nc} \partial u_{11}} & \cdots & \frac{\partial^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*)}{\partial u_{nc} \partial u_{nc}} \end{bmatrix}. \quad (22)$$

$\nabla^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*)$ is a diagonal matrix and its diagonal element is

$$\frac{\partial^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*)}{\partial u_{ik} \partial u_{ik}} = \frac{\beta}{\sqrt{c}} u_{ik}^{m-2}. \quad (23)$$

Since $u_{ik} > 0$, Hessian matrix $\nabla^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*)$ is positive definite. Having proved that \mathbf{U}^* is a stagnation point of $J_{\text{SSOD-AFW}}(\mathbf{U})$ and $\nabla^2 J_{\text{SSOD-AFW}}(\mathbf{U}^*)$ is positive definite, we conclude \mathbf{U}^* is the local minimum of $J_{\text{SSOD-AFW}}(\mathbf{U})$. Then we have $J_{\text{SSOD-AFW}}(\mathbf{U}^{(t)}) \geq J_{\text{SSOD-AFW}}(\mathbf{U}^*) = J_{\text{SSOD-AFW}}(\mathbf{U}^{(t+1)})$, where $\mathbf{U}^{(t)}$ is the membership matrix after the t th iteration in (19) and $\mathbf{U}^* = \mathbf{U}^{(t+1)}$ is the one after the $(t + 1)$ th iteration. Therefore, objective function $J_{\text{SSOD-AFW}}$ in (8) is nonincreasing by updating $\mathbf{U} = (u_{ik})_{n \times c}$ using formula (19). \square

Lemma 3. *Objective function $J_{\text{SSOD-AFW}}$ in (8) is nonincreasing by updating $\mathbf{V} = (v_{jk})_{N \times c}$ using (15).*

Proof. Similar to Lemma 2, when \mathbf{U} and \mathbf{O} are fixed, we just need to prove that the Hessian matrix of Lagrangian of $J_{\text{SSOD-AFW}}(\mathbf{V})$ at \mathbf{V}^* is positive definite, where \mathbf{V}^* is computed by (15). The Hessian matrix is denoted as $H(\mathbf{V}) = (h_{jk,lm}(\mathbf{V}))$, whose element is expressed as follows:

$$h_{jk,lm}(\mathbf{V}) = \frac{\partial^2 L_1(\mathbf{V})}{\partial v_{jk} \partial v_{lm}} = \begin{cases} q(q-1)v_{jk}^{q-2} \sum_{i=1}^n u_{ik}^m (x_{ij} - o_{kj})^2, & \text{if } j=l, k=m, \\ 0, & \text{otherwise.} \end{cases} \quad (24)$$

Since $q > 1$ and $v_{jk} > 0$, the diagonal entries of the diagonal matrix are apparently positive. Therefore, Hessian matrix $H(\mathbf{V})$ is positive definite. $J_{\text{SSOD-AFW}}(\mathbf{V})$ attains its local minimum at \mathbf{V}^* computed by (15). This completes the proof. \square

Lemma 4. *Objective function $J_{\text{SSOD-AFW}}$ in (8) is nonincreasing when $\mathbf{O} = (o_{kj})_{c \times N}$ is updated using (17).*

The proof of Lemma 4 is similar to Lemma 2.

Lemma 5. *Objective function $J_{\text{SSOD-AFW}}$ in (8) is bounded, there exists a constant B , and it satisfies $J_{\text{SSOD-AFW}} < B$.*

Proof. $\forall i \in \{1, \dots, n\}$ and $\forall k \in \{1, \dots, c\}$, we have $0 < u_{ik} \leq 1$. Thus, $u_{ik}^m \log u_{ik}^m - u_{ik}^m$ is monotonically decreasing with respect to u_{ik} and hence $-1 \leq u_{ik}^m \log u_{ik}^m - u_{ik}^m < 0$. Apparently, the first term and the second term of the objective function $J_{\text{SSOD-AFW}}$ are also bounded. In consequence, objective function $J_{\text{SSOD-AFW}}$ in (8) is bounded. \square

Theorem 6. *The objective function $J_{\text{SSOD-AFW}}$ converges to a local minimum by updating \mathbf{V} , \mathbf{O} , and \mathbf{U} using formulas (15), (17), and (19), which subjects to the constraints in (9).*

Proof. Lemmas 2, 3, and 4 verify that objective function in (8) is nonincreasing under iterations according to (15), (17), and (19). Lemma 5 shows that $J_{\text{SSOD-AFW}}$ has a finite bound. Though the parameter α_{ik} needs to be updated in the iteration process, it is a constant in the problem solving using Lagrangian multiplier technique. So α_{ik} does not affect the convergence of the SSOD-AFW algorithm. Combining the above conclusions, $J_{\text{SSOD-AFW}}$ is sure to converge to a local

minimum through iterations of \mathbf{V} , \mathbf{O} , and \mathbf{U} by (15), (17), and (19). \square

4. Experiments and Analysis

Comprehensive experiments and analysis on a synthetic dataset and several real-world datasets are conducted to show the effectiveness and superiority of the proposed SSOD-AFW. We compared the proposed algorithm with two the-state-of-the-art unsupervised outlier algorithms, LOF [15] and KNN [13], one supervised method SVDD [32], and one semisupervised method EODSP [23].

4.1. Evaluation Metrics. For numerically performance evaluation of outlier detection algorithms, three metrics, namely, accuracy [7], AUC [33], and rank-power [16], are used in this paper.

Let M be the number of true outliers that a dataset X contains and S denotes the number of true outliers detected by an algorithm. In experiments, top M most suspicious instances are detected out. Then the accuracy is given by

$$\text{Accuracy} = \frac{S}{M}. \quad (25)$$

The receiver operating characteristic (ROC) curve represents the trade-off relationship between the detection rate and the false alarm rate. In general, the area under the ROC curve (AUC) is used to measure the performance of outlier detection method, and the value of AUC for ideal detection performance is close to one.

For a given outlier detection algorithm, true outliers occupy top positions with respect to the nonoutliers among M suspicious instances; then the rank-power (RP) of the algorithm is said to be high. If S is the number of true outliers found within top M instances and R_i denotes the rank of the i th true outlier, then the metric rank-power (RP) is given by

$$\text{RP} = \frac{S(S+1)}{2 \sum_{i=1}^S R_i}. \quad (26)$$

RP reaches the maximum value 1 when all S true outliers are in the top S positions. Larger value of RP implies better performance of an algorithm.

4.2. Experiments on Synthetic Dataset. A two-dimensional synthetic dataset with two cluster patterns is generated from Gaussian distribution to intuitively compare the outlier detection results of the proposed method against the other four algorithms mentioned above. The mean vectors of the two clusters are $\mu_1 = (7.5, 9)^T$ and $\mu_2 = (1, 3)^T$, respectively, and the covariance matrixes of them are $\Sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$ and $\Sigma_2 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. As Figure 2(a) shows, a total of 199 samples are contained in the synthetic dataset, in which there are 183 normal samples (within two clusters) and 16 outliers (cluttered between two clusters). 13 normal objects are labeled and marked as symbol “ \times ” and 5 outliers are labeled and marked with symbol “ $*$ ”, while the rest samples are unlabeled marked with “ \cdot .” Figures 2(b)–2(f), respectively, illustrate the

outlier detection results on the synthetic dataset by using LOF, KNN, SVDD, EODSP, and SSOD-AFW, where the red colored symbols “ \circ ” denote the detected suspicious outliers. Here, the value of parameter k (size of neighborhood) in LOF and KNN is assigned to 3. Gauss kernel function is chosen in SVDD and we set the bandwidth $\sigma = 0.3$ and the trade-off coefficient $C = 0.45$. Besides, the Euclidean distance threshold T in EODSP is set as 0.1 and the percentage of negative set is set to $l = 10\%$. The parameter settings of the proposed algorithm are $m = 2.3$, $K = 0.85$, and $q = 2.23$. In addition to SVDD, the top 16 objects with the highest outlying scores are considered as the results in the other four algorithms.

In Figure 2, it is noticeable that the unsupervised methods LOF and KNN as well as the supervised SVDD fail to completely detect all of the 5 labeled outliers. Nevertheless, some normal points in clusters are badly misjudged as outliers. In contrast, the semisupervised EODSP algorithm and the proposed SSOD-AFW algorithm successfully detect all of the 5 labeled outliers. However, EODSP does not completely detect all the unlabeled true outliers, and several true normal samples are improperly identified as outliers. It is concluded from Figure 2 that the proposed algorithm finds all the true outliers in the synthetic dataset and excludes the normal samples, while the other methods do not.

Figure 3 numerically presents the performance evaluation of outlier detection using LOF, KNN, SVDD, EODSP, and SSOD-AFW for the synthetic dataset. From Figure 3 we see that the values of accuracy, AUC, and RP of the proposed algorithm all reach 1, outperforming the other methods.

Furthermore, during the experimental process shown in Figure 3, the feature weights of the synthetic dataset learned by formula (15) in our method are $v_{11} = 0.6727$, $v_{12} = 0.5985$, $v_{21} = 0.3273$, and $v_{22} = 0.4015$. To strengthen the effectiveness of feature weights in the proposed SSOD-AFW algorithm, a comparative analysis of the weighted and the nonweighted versions is implemented on the synthetic dataset, respectively. Considering the nonweighted scenario of the proposed algorithm, the outlier detection result on the synthetic dataset is presented in Figure 4. As can be observed from Figure 4, the nonweighted SSOD-AFW ends up tagging 15 true outlying and 1 normal samples as outliers, with one unlabeled true outlier missed.

4.3. Experiments on Real-World Datasets

4.3.1. Introduction of Datasets. For further verification of the effectiveness of the proposed algorithm, five real-world datasets from UCI Machine Learning Repository [34] (i.e., Iris, Abalone, Wine, Ecoli, and Breast Cancer Wisconsin (WDBC)) are employed to test the performance of the proposed algorithm against LOF, KNN, SVDD, and EODSP. As mentioned in Aggarwal and Yu [35], one way to test the performance of an outlier detection algorithm is to run it on the dataset and calculate the percentage of points belonging to the rare classes. So a small amount of samples from the same class are randomly selected as outlying objects or as target objects, for the five datasets. For instance, the original Iris dataset incorporates 150 objects with 50 objects in each

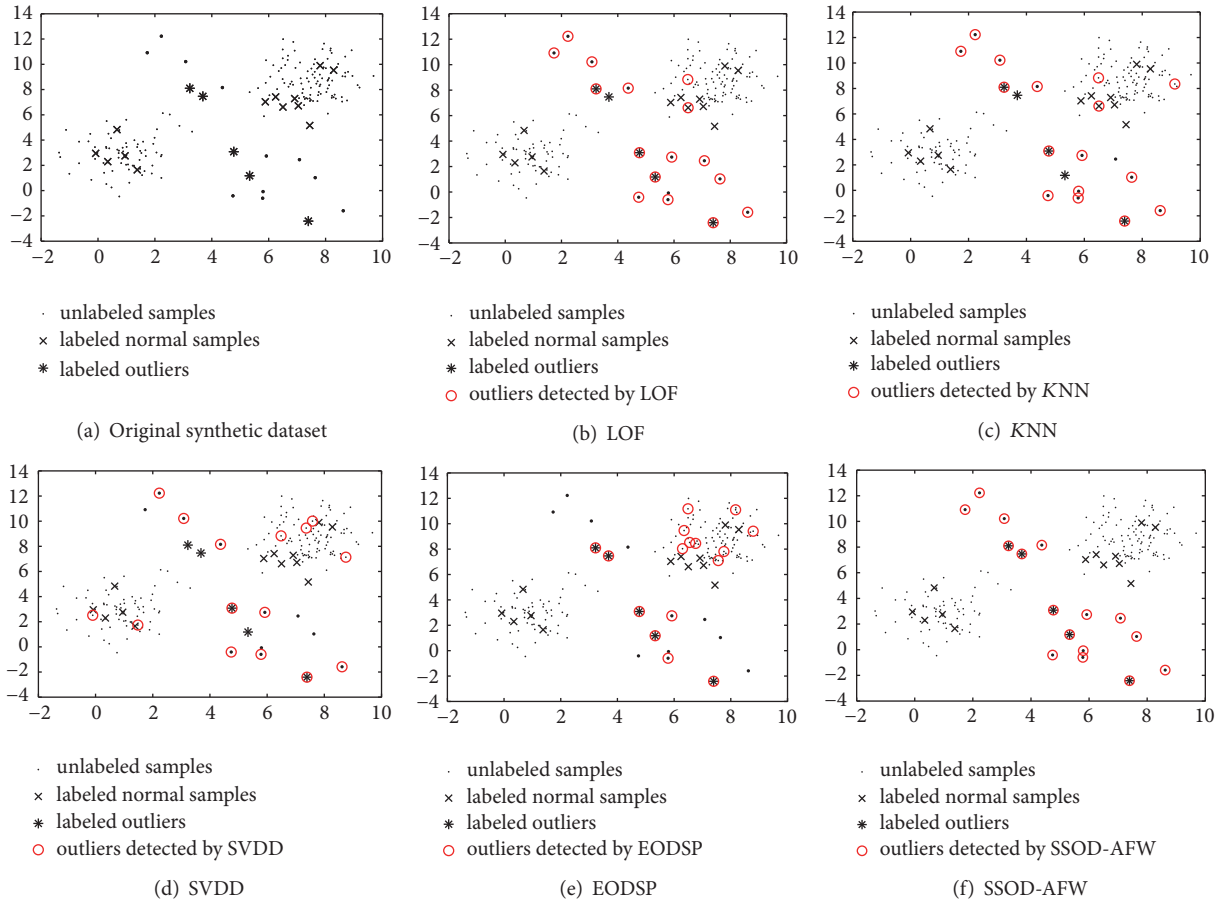


FIGURE 2: Outlier detection results of different algorithms on the two-dimensional synthetic dataset.

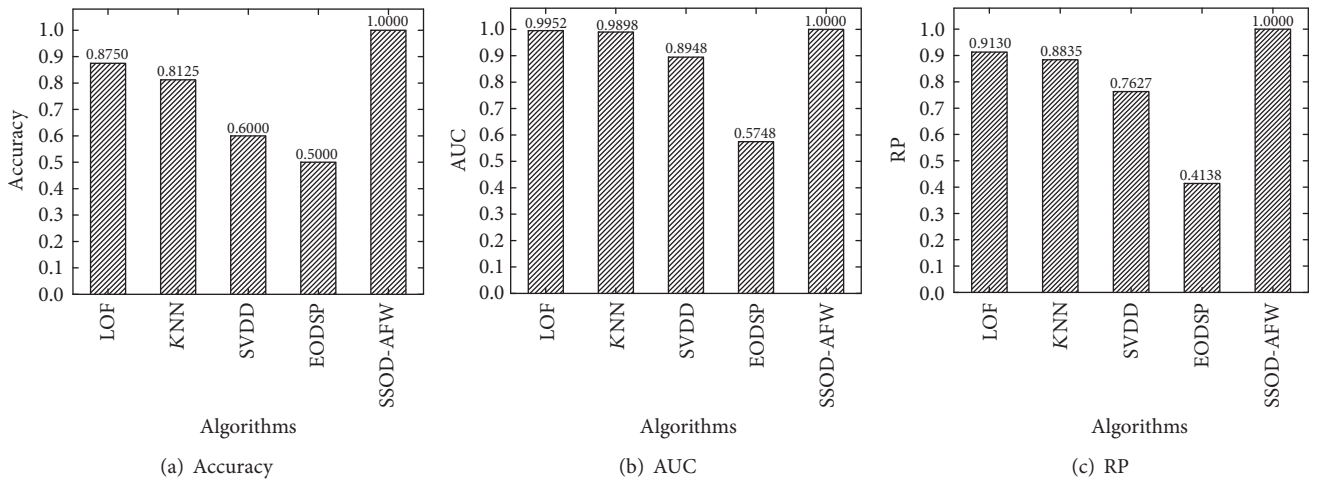


FIGURE 3: Performance comparison of different algorithms on the synthetic dataset.

of three classes. We randomly selected 26 objects from class “Iris-virginica” as target outliers and all objects in the other two classes are considered as normal objects. The other four datasets are similarly preprocessed and more detailed description about the five real-world datasets is given in Table 1.

4.3.2. Experimental Result Analysis. We compare the outlier detection performance of the proposed algorithm with LOF, KNN, SVDD, and EODSP on real-world datasets. Each method has its own parameters, and the detailed parameter settings of each algorithm are as follows. The parameters of the proposed algorithm are $m = 2.1$, $K = 0.65$, and $q = 1.53$

TABLE I: Description of real-world datasets.

Dataset	Instances	Features	Outlying classes	Outliers (percent)	Clusters	Prior information
Iris	126	4	“Virginica”	26 (20.63%)	2	10 labeled normal samples, 4 labeled outliers
Abalone	4177	8	“1”-“4,” “16”-“27,” “29”	335 (8.02%)	11	11 labeled normal samples, 18 labeled outliers
Wine	130	13	“3”	11 (8.46%)	2	9 labeled normal samples, 4 labeled outliers
Ecoli	336	9	“omL,” “imL,” “imS”	9 (2.68%)	5	11 labeled normal samples, 3 labeled outliers
WDBC	387	30	“Malignant”	30 (7.75%)	1	10 labeled normal samples, 8 labeled outliers

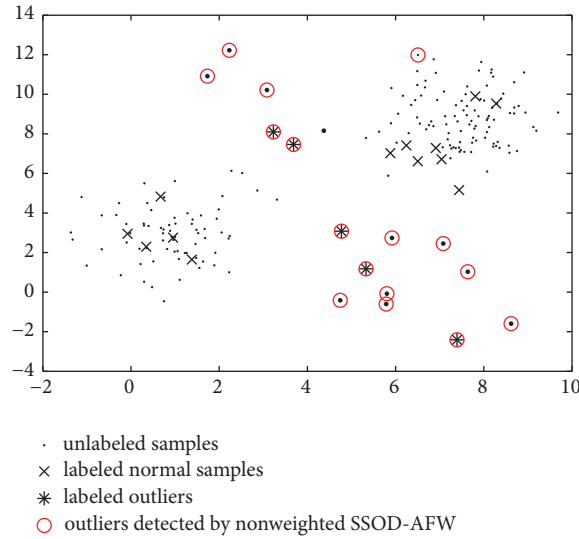


FIGURE 4: Outlier detection result of the nonweighted SSOD-AFW on the synthetic dataset.

for all the five datasets. The strategy of parameter selection for SSOD-AFW will be discussed in the later subsection called parameter analysis. For the other algorithms, those parameters are set exactly as mentioned in their references. It is well known that LOF and KNN have high dependency on the neighborhood parameter k . In this paper we set $k = 5$ for datasets Iris and WDBC, $k = 3$ for dataset Abalone, $k = 10$ for dataset Wine, and $k = 10$ for dataset Ecoli. For SVDD method, Gaussian kernel function is employed and the bandwidths $\sigma = 0.45$ and $C = 0.5$ on all of the five real-world datasets. In EODSP, the Euclidean distance threshold T is set as 0.1 and the percentage of negative set l is set as 30% for Iris and Abalone datasets, and $T = 0.5$, $l = 30\%$ for datasets Ecoli, Wine, and WDBC. Since we randomly select outliers from target classes for each dataset, each experiment is repeated 10 times with the same number of different outliers. The average accuracy, AUC, and RP are calculated as the criteria of performance of various detection methods.

Figure 5 illustrates the outlier detection results of SSOD-AFW algorithm against LOF, KNN, SVDD, and EODSP, respectively, on the five real-world datasets. As can be seen from Figure 5, the proposed algorithm can accurately identify outliers according to the cluster structure of a dataset, with the guidance of the label knowledge. It shows distinct superiority over the other unsupervised (LOF, KNN), semisupervised (EODSP), and supervised (SVDD) methods. In particular, the outlier detection accuracy of SSOD-AFW in Figure 5(a) is significantly higher than the others, especially

for datasets Iris and Wine. One can know from Figure 5(b) that the AUC values of our method are always higher than the others for all datasets except for WDBC. In terms of RP, SSOD-AFW performs better than the other four algorithms on datasets Iris and Wine, whereas slightly poorer than SVDD on Abalone, poorer than LOF on Ecoli, and poorer than KNN on WDBC, seen as in Figure 5(c).

It is worth mentioning that the experiment of the proposed algorithm on WDBC involves one-class clustering problem. Although one-class clustering task is generally meaningless, one-class clustering-based outlier detection is especially meaningful and feasible in our proposal because our approach does not require that the membership degrees must sum up to 1. This is one of the powerful and important characteristics of the proposed algorithm.

4.3.3. The Influence of the Proportion of Labeled Data on Outlier Detection. In this subsection, we will investigate the influence of the proportion of labeled samples on the accuracies of our methodology. Two typical situations are considered and tested. The first one is that the proportion of labeled outliers increases when the number of labeled normal objects is fixed at a certain constant. The other one is that the percent of labeled normal samples varies while the quantity of labeled outliers is fixed. So two groups of experiments are designed to compare the accuracies of the proposed algorithm against the EODSP, in the situations of different

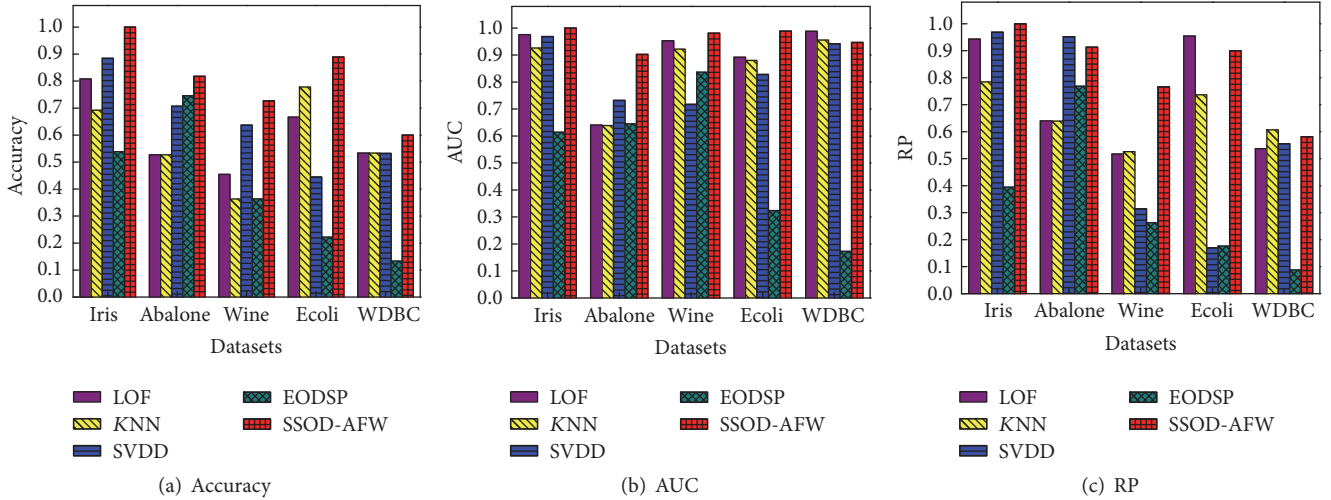


FIGURE 5: Performance comparison of various algorithms on the real-world datasets.

percent of labeled outliers and normal samples, respectively, on the datasets Iris, Abalone, Wine, Ecoli, and WDBC. In the two experiments, the percent of labeled outliers or labeled normal samples ranges from 0% to 40%, respectively, when the number of another kind of labeled objects is fixed. We randomly select a certain number of labeled outliers or normal samples from each dataset, each experiment is repeated 10 times, and the average accuracies of SSOD-AFW and EODSP are computed.

Figure 6 shows results of the first group of experiments where the percent of labeled outliers varies from 0% to 40%. One can see from Figure 6 that the accuracies of the two semisupervised algorithms are roughly increased with the proportion of labeled outliers becoming larger. This powerfully supports the idea that semisupervised outlier detection algorithms can improve the accuracy of outlier detection by using prior information. Furthermore, the SSOD-AFW achieves a better accuracy than EODSP algorithm for the same proportion of labeled outliers on the five datasets. Especially for Wine, the accuracy of SSOD-AFW is 40% higher than that of EODSP. EODSP addresses the problem of detecting outliers with only few labeled outliers as training data. The labeled normal instances are extracted according to the maximum entropy principle, where the entropy is computed only using the distance between each testing sample and all the labeled outliers. That makes EODSP not flexible as our proposed method due to information deficiencies.

Figure 7 illustrates the accuracy comparison of the proposed algorithm and EODSP, when the proportion of labeled normal samples increases from 0% to 40% and the percent of labeled outliers is fixed. Note that our method obtains a better accuracy than EODSP on all of the five real-world datasets. The accuracy of the proposed algorithm gets larger when the percent of labeled normal samples increases. As mentioned, EODSP emphasizes the semisupervised outlier detection only with few labeled outliers in the initial dataset, but without considering any labeled normal objects. Therefore,

the accuracy of EODSP algorithm keeps harper with various proportions of labeled normal objects and always equals the accuracy value of 0% labeled normal samples as well.

4.3.4. Parameter Analysis. The parameters m , K , and q are important in our proposed algorithm, which affect the performance of SSOD-AFW. In this section, the influence of each parameter on outlier detection accuracy is studied.

The parameter m is the fuzzification coefficient. Figure 8(a) analyzes the relationship between the outlier detection accuracy of our proposed algorithm and parameter m , with m varying from 1.5 to 5.0. The results imply that the highest accuracy is achieved when m ranges in $[2, 4]$. So it is reasonable that m value in the experiments shown in Figure 5 has been set as 2.1. The parameter $K \in (0, 1)$ controls the importance of the label information in the result of outlier detection. Outlier detection accuracies are testified by varying K from 0.1 to 0.9, which are shown in Figure 8(b). The overall tendency is that the accuracies become larger as K increases. The best results of the proposed algorithm occur and keep stable when $K \geq 0.7$. Finally, from Figure 8(c), we conclude that the feature weight index q has small influence on the accuracy of SSOD-AFW in the situation that the other parameters maintain the same settings. So the proposed algorithm is not sensitive to the parameter q . In general, the parameter q is suggested to select a constant from $(1, 4]$.

4.3.5. Execution Time Analysis. Figure 9 analyzes the average running time of the proposed algorithm against the other algorithms on five real-world datasets. The experimental environment is Windows XP systems, MATLAB 7.1 platform, 3 GHz CPU, 2 GB RAM. Because the volume of Abalone dataset is far greater than the other four datasets, the running times of various datasets are distinctly different. In order to facilitate the display, in Figure 9 the horizontal coordinate axis is translated downward a short distance. The result indicates that the proposed algorithm is more time-saving than the

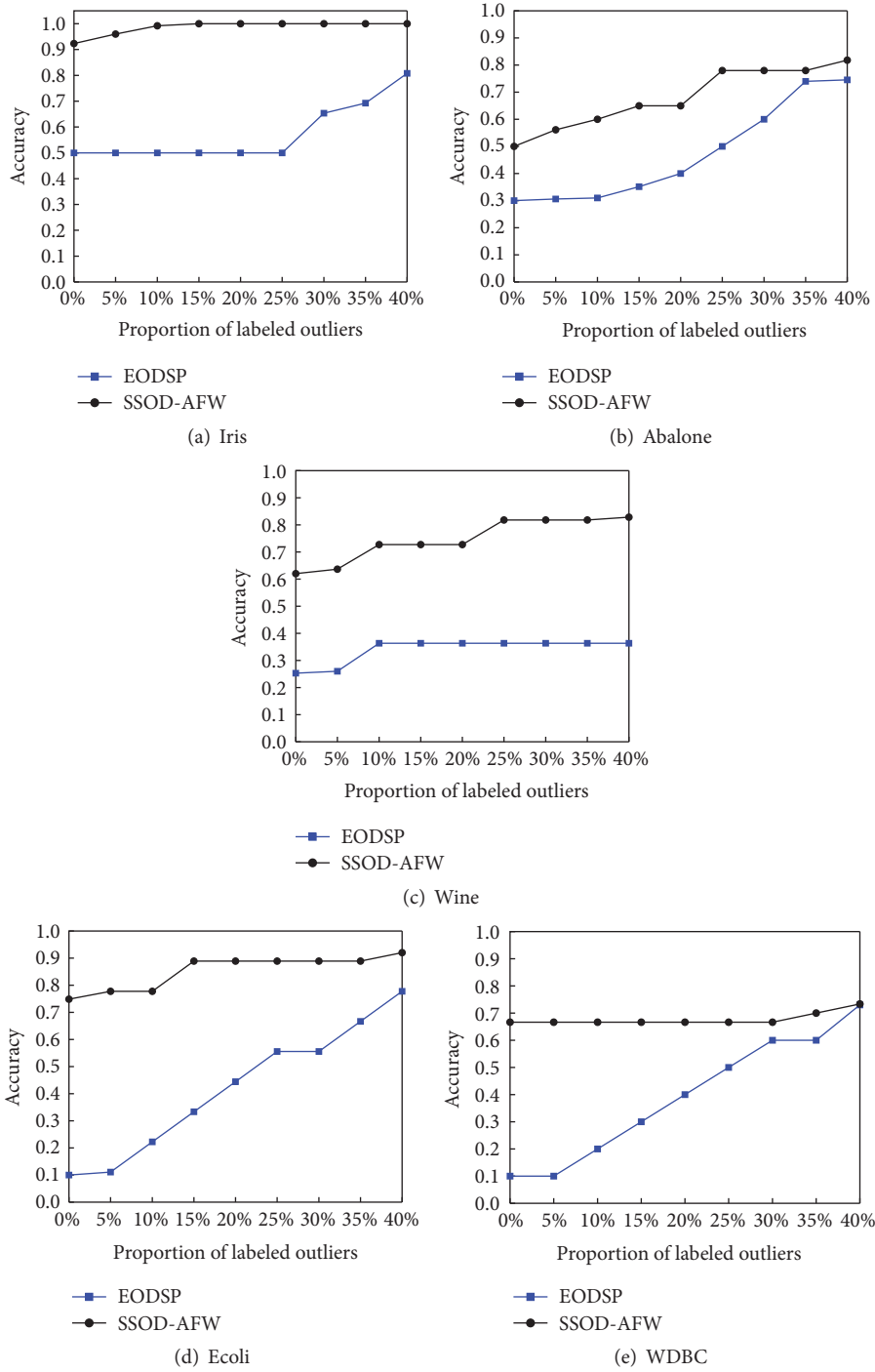


FIGURE 6: Accuracy analysis of algorithms EODSP and SSOD-AFW with different percent of labeled outliers on the real-world datasets.

other four typical outlier detection algorithms, except for KNN on dataset Wine. In whole, the execution time of the SSOD-AFW is comparable to that of KNN and less than those of other algorithms on most of the datasets.

5. Conclusions

In order to detect outliers more precisely, a semisupervised outlier detection algorithm based on adaptive feature

weighted clustering, called SSOD-AFW, is proposed in this paper. Distinct weights of each feature with respect to different clusters are considered and obtained by adaptive iteration, so that the negative effects of irrelevant features on outlier detection are weakened. Moreover, the proposed method makes full use of the prior knowledge contained in datasets and detects outliers in virtue of the cluster structure. It is verified by a series of experiments that the proposed SSOD-AFW algorithm is superior to other typical unsupervised,

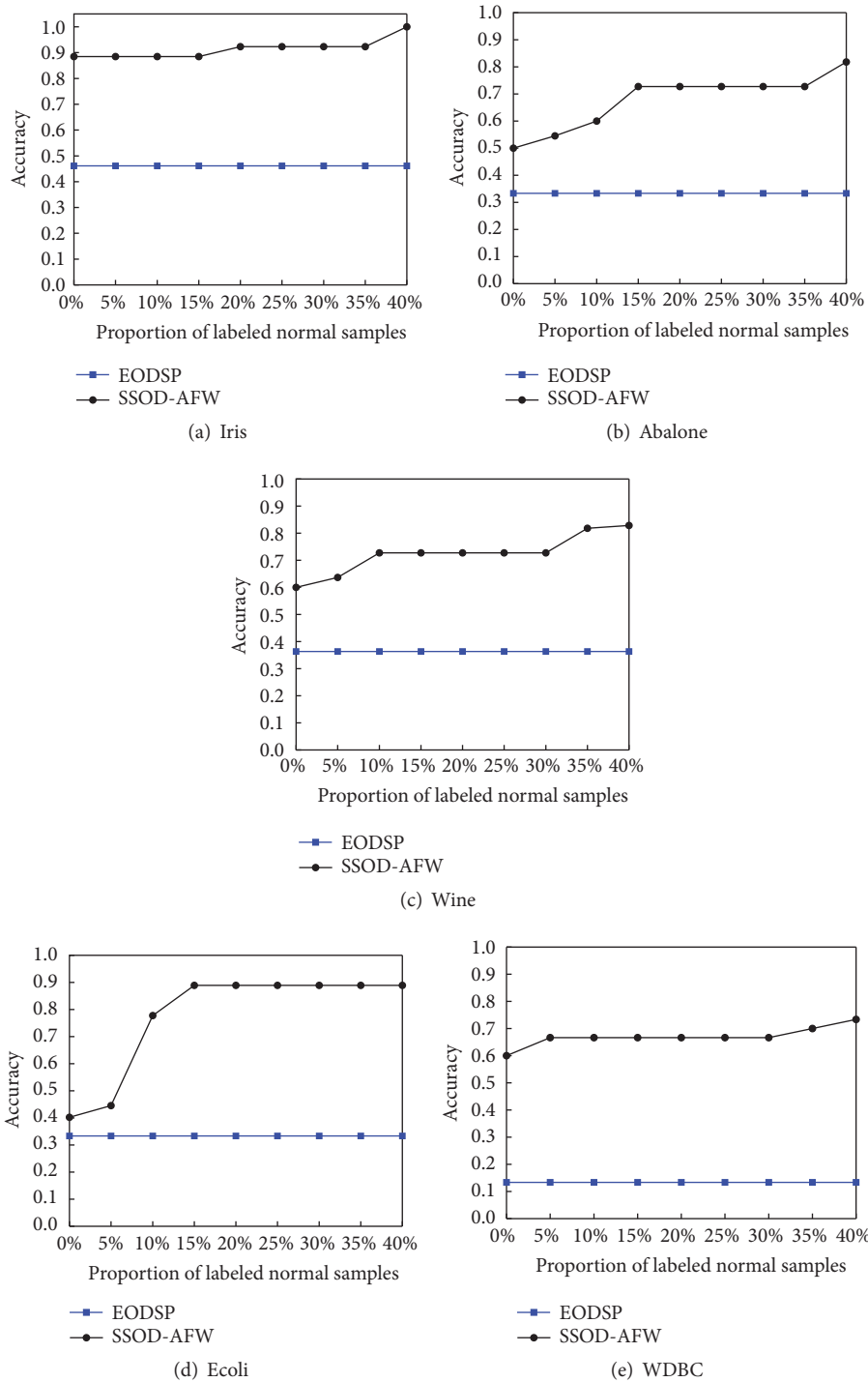


FIGURE 7: Accuracy analysis of algorithms EODSP and SSOD-AFW with different percent of labeled normal samples on the real-world datasets.

semisupervised, and supervised algorithms in both outlier detection precision and running speed.

In this paper, we present a new semisupervised outlier detection method which utilizes labels of a small amount of objects. However, our method assumes that the labels

of objects are reliable and does not consider mislabel punishment in the new objective function. Therefore, a robust version of the proposed method dealing with noisy or imperfect labels of objects deserves further studies. Moreover, since only one typical dissimilarity measure named Euclidean

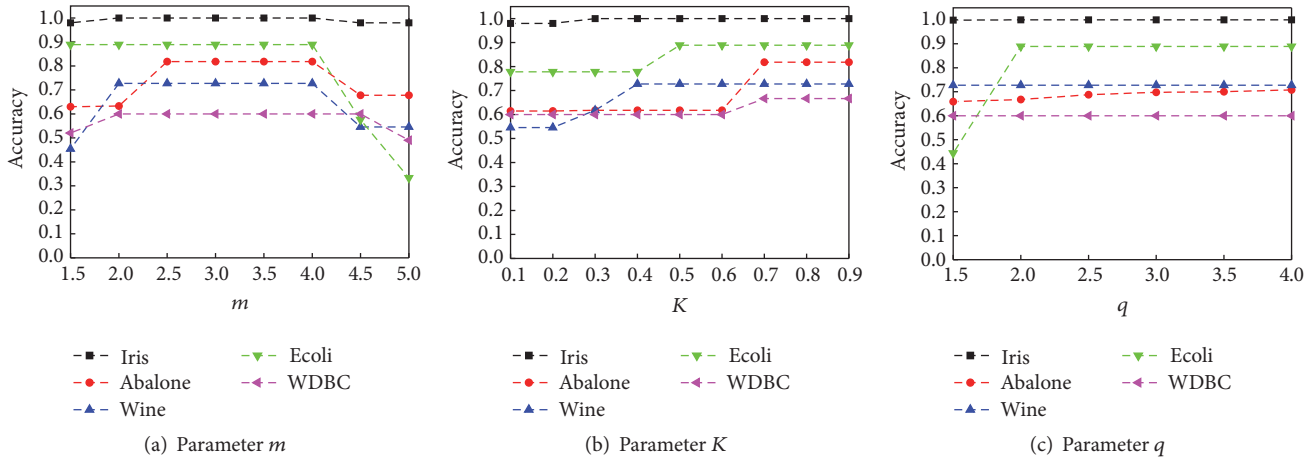


FIGURE 8: Outlier detection accuracy of the proposed algorithm under various parameters on the real-world datasets.

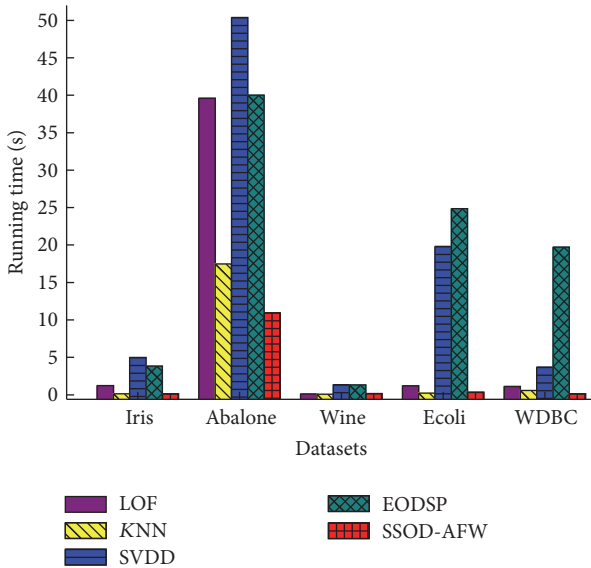


FIGURE 9: Execution time comparison of different algorithms on the real-world datasets.

distance is discussed in our method, the SSOD-AFW algorithm is limited to outlier detection for numerical data. The future research aims at extending our method to mixed-attribute data in more real-life applications such as fault diagnosis in industrial process or network anomaly detection.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

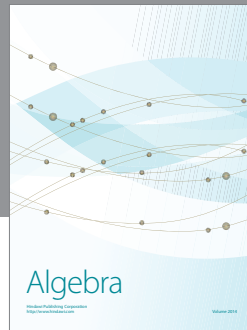
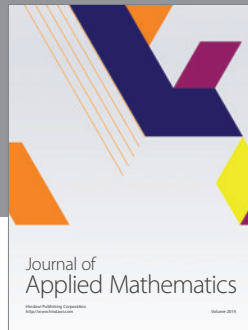
Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant no. 11471001).

References

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Elsevier, 2011.
- [2] D. M. Hawkins, *Identification of Outliers*, Chapman & Hall, London, UK, 1980.
- [3] V. Barnett and T. Lewis, *Outliers in Statistical Data*, John Wiley & Sons, Chichester, UK, 1994.
- [4] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, 2004.
- [5] B. Sheng, Q. Li, W. Mao, and W. Jin, "Outlier detection in sensor networks," in *Proceedings of the 8th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc '07)*, pp. 219–228, Montréal, Canada, September 2007.
- [6] A. P. James and S. Dimitrijević, "Inter-image outliers and their application to image classification," *Pattern Recognition*, vol. 43, no. 12, pp. 4101–4112, 2010.
- [7] J. Huang, Q. Zhu, L. Yang, and J. Feng, "A non-parameter outlier detection algorithm based on Natural Neighbor," *Knowledge-Based Systems*, vol. 92, pp. 71–77, 2016.
- [8] J. M. Shepherd and S. J. Burian, "Detection of urban-induced rainfall anomalies in a major coastal city," *Earth Interactions*, vol. 7, no. 4, pp. 1–17, 2003.
- [9] O. Alan and C. Catal, "Thresholds based outlier detection approach for mining class outliers: an empirical case study on software measurement datasets," *Expert Systems with Applications*, vol. 38, no. 4, pp. 3440–3445, 2011.
- [10] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [11] D. Tax, A. Ypma, and R. Duin, "Support vector data description applied to machine vibration analysis," in *Proceedings of the 5th Annual Conference of the Advanced School for Computing and Imaging*, pp. 398–405, Heijen, The Netherlands, June 1999.
- [12] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.
- [13] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," *SIGMOD Record (ACM Special Interest Group on Management of Data)*, vol. 29, no. 2, pp. 427–438, 2000.

- [14] E. M. Knox and R. T. Ng, "Algorithms for mining distance based outliers in large dataset," in *Proceedings of the International Conference on Very Large Data Bases*, pp. 392–403, Citeseer, New York, NY, USA, 1998.
- [15] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," *ACM SIGMOD Record*, vol. 29, no. 2, pp. 93–104, 2000.
- [16] J. Ha, S. Seok, and J.-S. Lee, "A precise ranking method for outlier detection," *Information Sciences*, vol. 324, pp. 88–107, 2015.
- [17] R. N. Dave, "Characterization and detection of noise in clustering," *Pattern Recognition Letters*, vol. 12, no. 11, pp. 657–664, 1991.
- [18] R. Smith, A. Bivens, M. Embrechts, C. Palagiri, and B. Szymanski, "Clustering approaches for anomaly based intrusion detection," *Proceedings of Intelligent Engineering Systems Through Artificial Neural Networks*, pp. 579–584, 2002.
- [19] Y. Shi and L. Zhang, "COID: a cluster-outlier iterative detection approach to multi-dimensional data analysis," *Knowledge and Information Systems*, vol. 28, no. 3, pp. 709–733, 2011.
- [20] J. Zhao, K. Liu, W. Wang, and Y. Liu, "Adaptive fuzzy clustering based anomaly data detection in energy system of steel industry," *Information Sciences*, vol. 259, pp. 335–345, 2014.
- [21] F. Angiulli and C. Pizzuti, "Fast outlier detection in high dimensional spaces," in *Proceedings of the 6th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD '02)*, vol. 2, pp. 15–26, August 2002.
- [22] M. Radovanović, A. Nanopoulos, and M. Ivanović, "Reverse nearest neighbors in unsupervised distance-based outlier detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 5, pp. 1369–1382, 2015.
- [23] A. Daneshpazhouh and A. Sami, "Entropy-based outlier detection using semi-supervised approach with few positive examples," *Pattern Recognition Letters*, vol. 49, pp. 77–84, 2014.
- [24] J. Gao, H. Cheng, and P.-N. Tan, "Semi-supervised outlier detection," in *Proceedings of the ACM Symposium on Applied Computing*, pp. 635–636, ACM, Dijon, France, April 2006.
- [25] Z. Xue, Y. Shang, and A. Feng, "Semi-supervised outlier detection based on fuzzy rough C-means clustering," *Mathematics and Computers in Simulation*, vol. 80, no. 9, pp. 1911–1921, 2010.
- [26] J. Z. Huang, M. K. Ng, H. Rong, and Z. Li, "Automated variable weighting in k-means type clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 657–668, 2005.
- [27] H. Frigui and O. Nasraoui, "Unsupervised learning of prototypes and attribute weights," *Pattern Recognition*, vol. 37, no. 3, pp. 567–581, 2004.
- [28] J. Zhou, L. Chen, C. P. Chen, Y. Zhang, and H. Li, "Fuzzy clustering with the entropy of attribute weights," *Neurocomputing*, vol. 198, pp. 125–134, 2016.
- [29] M. Hassan, A. Chaudhry, A. Khan, and M. A. Iftikhar, "Robust information gain based fuzzy c-means clustering and classification of carotid artery ultrasound images," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 593–609, 2014.
- [30] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, 1993.
- [31] M.-S. Yang and K.-L. Wu, "Unsupervised possibilistic clustering," *Pattern Recognition*, vol. 39, no. 1, pp. 5–21, 2006.
- [32] S. M. Guo, L. C. Chen, and J. S. H. Tsai, "A boundary method for outlier detection based on support vector domain description," *Pattern Recognition*, vol. 42, no. 1, pp. 77–83, 2009.
- [33] T. Fawcett, "Roc graphs: notes and practical considerations for researchers," *Machine Learning*, vol. 31, no. 1, pp. 1–38, 2004.
- [34] C. Blake and C. J. Merz, "UCI repository of machine learning databases," 1998.
- [35] C. C. Aggarwal and P. S. Yu, "Outlier detection for high dimensional data," in *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '01)*, Santa Barbara, Calif, USA, May 2001.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

