Applied Soft Computing 134 (2023) 109995

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Anomaly detection based on weighted fuzzy-rough density

Zhong Yuan [a], Baiyang Chen [a], Jia Liu [b], Hongmei Chen [c], Dezhong Peng [a], Peilin Li [d,e,f,*]

[a] College of Computer Science, Sichuan University, Chengdu 610065, China
[b] School of Computer and Software Engineering, Xihua University, Chengdu 610039, China
[c] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China
[d] Dept. of Orthodontics, West China Hospital of Stomatology, Sichuan University, Chengdu 610065, China
[e] National Clinical Research Center for Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu 610065, China
[f] State Key Laboratory of Oral Diseases, West China Hospital of Stomatology, Sichuan University, Chengdu 610065, China

ARTICLE INFO

Article history:
Received 24 September 2022
Received in revised form 13 December 2022
Accepted 2 January 2023
Available online 7 January 2023

Keywords:
Anomaly detection
Granular computing
Fuzzy rough set theory
Weighted density
Mixed data

ABSTRACT

The density-based method is a more widely used anomaly detection. However, most of the existing density-based methods mainly focus on dealing with certainty data and do not consider the problem of uncertainty and fuzziness of the data. Fuzzy rough set theory, as an important mathematical model of granular computing, provides an effective method for information processing of uncertain data. For this reason, this paper proposes an anomaly detection based on fuzzy-rough density. First, the fuzzy-rough density is defined to describe the degree of aggregation of objects. Then, fuzzy entropy is introduced to compute the weights of each attribute. Further, an anomaly score is constructed to characterize the anomaly degree of the samples, which takes into account both the density and fuzziness of the samples. Finally, extensive experiments are conducted on publicly available data with nine popular detection methods. The experimental results show that the proposed method achieves better performance on three types of datasets.

1. Introduction

An anomaly (aka outlier) is defined as a data sample (or object) that is significantly different from the rest of the data distribution. Since the presence of anomalies may affect the results of data analysis, removing them becomes a critical pre-processing step in some data analysis models, such as clustering analysis [1,2]. However, Anomaly Detection (AD) has a wide range of applications in many real-life fields, such as industrial processes [3], power system maintenance [4], dangerous driving detection [5], fraud detection [6], medical anomaly detection [7], etc. These applications all require high detection rates and low false alarm rates for AD algorithms. Therefore, it becomes especially important to further propose AD methods with good performance.

Most AD methods aim to find outliers from unlabeled data. Usually, they can be classified into statistical [8,9], distance [10,11], density [12,13], and clustering [14,15] methods depending on the techniques used in the model. Among them, density-based methods are the older known methods for AD. Recently, such methods have attracted extensive research. They are based on the core idea that outliers usually appear in low-density regions, while non-outliers (aka inliers) are assumed to appear in dense neighborhoods. The Local Outlier Factor (LOF) method [12] is the first density-based method that has been proposed. Its purpose is to address the inadequacy of existing work that treats outliers as binary classifications. Based on the idea of the LOF algorithm, density-based detection methods have been widely studied [13,16–20]. In most of these models, Euclidean distance is used to define the anomaly scores. Therefore, they may not be suitable for detecting outliers in mixed attribute data. In addition, they usually use a deterministic approach to construct detection models, which may not reflect information such as the granular knowledge structure, uncertainty, and fuzziness of the data.

The fuzzy-rough computing model is an important granular computing model that has attracted a lot of attention. Dubois and Prade [21] proposed the Fuzzy Rough Set (FRS) model by combining the advantages of fuzzy sets and rough sets for the first time. After that, many researchers have done a lot of research on the extension of its model, such as kernelized FRSs [22], fuzzy neighborhood rough sets [23], covering-based variable precision FRSs [24], probability granular distance-based fuzzy rough set [25], etc. In addition, the fuzzy-rough computing model has been successfully applied in many directions, such as attribute reduction [26,27], multi-attribute decision-making [28–32], fuzzy-rough clustering [33], rule induction [34], etc. However, there

* Corresponding author at: Dept. of Orthodontics, West China Hospital of Stomatology, Sichuan University, Chengdu 610065, China.
E-mail addresses: yuanzhong@scu.edu.cn (Z. Yuan), farstars@qq.com (B. Chen), jialiu@mail.xhu.edu.cn (J. Liu), hmchen@swjtu.edu.cn (H. Chen), pengdz@scu.edu.cn (D. Peng), lipeilin@scu.edu.cn (P. Li).

https://doi.org/10.1016/j.asoc.2023.109995
1568-4946/

has been little work on fuzzy-rough computing models for finding outliers [35,36]. The potential of FRS theory in the field of anomaly detection has not been deeply explored.

In this paper, we propose a fuzzy-rough density-based anomaly detection, which takes into account both the density and uncertainty information of the samples. First, the concept of density is introduced into FR theory, thus the fuzzy-rough density is defined to characterize the isolation or irregularity of samples based on fuzzy relations. The smaller the fuzzy-rough density around a sample, the more irregular it is and the more likely it is to be an anomaly. Then, fuzzy entropy is introduced to construct weights that characterize the density of samples under each attribute. Further, a score characterizing the degree of abnormality of the sample is constructed by fusing fuzzy-rough density and weights. Finally, a simple and effective detection algorithm is given, namely, the Weighted Fuzzy-Rough Density-based Anomaly (WFRDA) detection algorithm. The analysis of experimental results on publicly available datasets demonstrates the superiority of the proposed WFRDA algorithm and its applicability to mixed data.

The remainder of this work is organized as follows. In the next section, the current status of research on density-based anomaly detection methods is briefly reviewed. In the third section, we present a piece of knowledge about FRS theory used in this paper. The fourth section proposes a weighted fuzzy-rough density-based anomaly detection. The analysis of experimental results is shown in the fifth section. Finally, a summary is given in the last section.

2. Related work

In response to the shortcomings of existing detection methods that treat anomalies as binary in nature, Breunig et al. [12] first proposed a density-based approach. They argued that for many cases it makes more sense to assign a LOF to each sample. The larger the LOF value of a sample, the more likely it is to be an anomaly. Since the LOF method only considers the difference between the density of a sample and its nearest neighbors, its effectiveness decreases when the density of an outlier is close to that of its neighbors. In view of this, Tang et al. [16] introduced a Connectivity-based Outlier Factor (COF) algorithm for finding anomalies, and demonstrated the improvement in validity and capability of the COF algorithm relative to the LOF algorithm through empirical analysis. Papadimitriou et al. [17] proposed a new method for evaluating outliers called Local Correlation Integral (LOCI). This method selects a sample as an outlier when the multigranularity deviation factor deviates from its standard deviation by a factor of three in its neighborhood. Jin et al. [18] designed a simple and effective local outlier detection method based on symmetric neighborhood relations. In estimating the sample density distribution, the method considers the nearest and reverse nearest neighbors of the samples and assigns the INFLuenced Outlierness (INFLO) to each object. Kriegel et al. [19] presented a new Local Outlier Probability (LoOP) outlier detection model. The model combines the idea of density-based local outlier scores with a probability statistics-oriented approach. Tang and He [13] proposed an anomaly detection method based on local kernel density estimation. The method considers three kinds of nearest neighbors in the local kernel density estimation: k-nearest neighbors, reverse nearest neighbors, and shared nearest neighbors. Li et al. [20] proposed a robust Directed density ratio Changing Rate-based Outlier Detection (DCROD) method. The method combines kernel density estimation with an extended neighbor set to calculate the local density of a sample.

The density-based methods described above can effectively detect outliers in most cases, but cannot effectively handle data containing uncertainty or fuzzy information. FRS theory is based on fuzzy-rough granules to form data models such as upper and lower approximation, dependency, and fuzzy information entropy, which can effectively deal with data containing uncertain or fuzzy information.

3. FRS theory

In fuzzy-rough computing, a data table without decisions is described by a two-tuple $DT = \langle OB, AT\rangle$. Herein, $OB = \{o_1, o_2, \ldots, o_n\}$ is a non-empty finite set of objects; AT is a non-empty finite set of conditional attributes; for any $o \in OB$ and $a \in AT$, $a(o)$ denotes the value of o with respect to attribute a.

Definition 1. Given $A \subseteq AT$, the fuzzy relation R_A with respect to A on OB is defined as

$$R_A : OB \times OB \to [0, 1]. \tag{1}$$

For any $o, p \in OB$, if R_A satisfies (1) reflexivity ($R_A(o, o) = 1$) and (2) symmetry ($R_A(o, p) = R_A(p, o)$), then R_A is a fuzzy similarity relation. $R_A(o, p)$ denotes the similarity between o and p and $R_A(o, p) = \min_{a\in A} R_a(o, p)$.

Definition 2. The fuzzy-rough granular structure induced by the fuzzy similarity relation R_A is defined as

$$G(R_A) = \{[o_1]_A, [o_2]_A, \ldots, [o_n]_A\}, \tag{2}$$

where $[o_i]_A = (R_A(o_i, o_1), R_A(o_i, o_2), \ldots, R_A(o_i, o_n))$. The cardinality of $[o_i]_A$ is $|[o_i]_A| = \sum_{j=1}^{n} R_A(o_i, o_j)$.

To better handle the data, Dubois and Prade proposed the following concept of FRSs [21].

Definition 3. For any fuzzy set $\tilde{O}$, the lower and upper approximations of $\tilde{O}$ are respectively defined as

$$\underline{R}_A\tilde{O}(o) = \inf_{p\in OB} \max\{(1 - R_A(o, p)), \tilde{O}(p)\}; \tag{3}$$

$$\overline{R}_A\tilde{O}(o) = \sup_{p\in OB} \min\{R_A(o, p), \tilde{O}(p)\}. \tag{4}$$

FRS model combines the advantages of fuzzy and rough sets and can effectively handle mixed attribute data containing fuzzy or imprecise information. Among them, the fuzzy-rough granule is the basic information granule of FRS theory, which can be used for the construction of anomaly detection models of uncertain data. In addition, when the fuzzy relations degenerate to equivalence relations, the fuzzy-rough granules become rough granules. This makes the fuzzy-rough computing model also applicable to handle rough and fuzzy information. Next, a density-based anomaly detection model is constructed based on fuzzy-rough granules.

4. A weighted fuzzy-rough density-based anomaly

In this section, we first construct an anomaly detection model; then an example is used to illustrate the proposed model; finally, the corresponding detection algorithm is designed and its time complexity is analyzed.

4.1. Detection model

For a given dataset to be detected, intuitive anomalies are those points that have highly irregular or uncommon values. The lower the frequency of occurrence of a sample value on each attribute, the more likely it is to be an anomaly. In addition, the "ideal" anomalies in the dataset are the sample values that are very irregular or rarely occur for each attribute. The sparsity of

a sample can be measured by calculating the average density through fuzzy-rough granules. The concept of average density is given in [37,38], however, it is only applicable to categorical attributes. In order to handle mixed data efficiently, we further define the fuzzy-rough density of a sample as follows.

Definition 4. The fuzzy-rough density of o with respect to A in OB is defined as

$$FRD_A(o) = \frac{\sum_{a\in A} Den_a(o)}{|A|}, \tag{5}$$

where $Den_a(o) = \sum_{p\in OB}[o]_a(p)/|OB|$ denotes the density of o on a.

If $FRD_A(o)$ is smaller, the number of neighbors around o is smaller in Definition 4. In other words, a smaller value of $FRD_A(o)$ means that o is more likely to be an anomaly.

In the above definition, all attributes have equal weights. Obviously, this is not reasonable. In some extreme cases, only a few attributes can determine whether a sample is an anomaly or not. That is, in practical applications, different attributes often contribute differently to the overall structure of the data. In fuzzy-rough computing, Hu et al. [39] introduced a fuzzy entropy and used it to measure uncertainties such as fuzziness and inconsistency. Since the purpose of anomaly detection is to detect rare samples that behave in unexpected ways or have anomalous properties. Uncertainty can be considered an anomalous property. Therefore, fuzzy entropy can be used to construct a detection model.

Definition 5. The fuzzy entropy with respect to R_A is computed by

$$FE(A) = -\frac{1}{|OB|}\sum_{o\in OB}\log_2\frac{|[o]_A|}{|OB|}. \tag{6}$$

According to Definition 5, for $a \in AT$, the larger the value of $FE(a)$, the more disorderly the distribution of its value domain is. When the value domain of a is uniformly distributed, the fuzzy entropy is the largest. It means it has the maximum uncertainty for attribute a. According to the above idea, the weighted fuzzy-rough anomaly score is constructed by combining fuzzy-rough density and fuzzy entropy.

Definition 6. The weighted fuzzy-rough density-based anomaly score of o is defined as

$$AS(o) = \sum_{a\in AT} W(a)(1 - FRD_a(o)), \tag{7}$$

where weight

$$W(a) = \frac{FE(a)}{\sum_{l=1}^{|AT|}(FE(a_l))} \in [0, 1]. \tag{8}$$

In the above definition, $W(a)$ uses fuzzy entropy to measure the distribution of the value domain over each attribute. Assuming the existence of a uniformly distributed value domain for each attribute, a certain attribute containing the maximum uncertainty will provide more anomalous features. Therefore, we should give more attention to this attribute and a larger weight should be given to it in the anomaly score. In addition, in the above definition, the weights of each attribute are normalized to [0, 1] and the sum of all values is 1. Furthermore, since both the density of the sample and the weight function are in the range of [0, 1], it is easy to verify that the range of the anomaly score is [0, 1].

According to the above idea, the weighted fuzzy-rough anomaly score can be used as an indicator to determine whether a sample is an anomaly, i.e., the larger the anomaly score of o, the higher the possibility that o is an anomaly.

Table 1
A mixed data table.

OB	a_1	a_2	a_3
o_1	A	0.2	6
o_2	A	0.1	2
o_3	B	0.4	3
o_4	C	0.1	5
o_5	C	0.4	1

Table 2
Corresponding fuzzy-rough densities.

$FRD_a(o)$	a_1	a_2	a_3
o_1	0.40	0.33	0.32
o_2	0.40	0.47	0.44
o_3	0.20	0.40	0.40
o_4	0.40	0.47	0.36
o_5	0.40	0.40	0.36

Definition 7. Given a threshold μ. For any $o \in OB$, if $AS(o) > \mu$, then o is called a weighted fuzzy-rough density-based anomaly in OB.

4.2. Detection example

In this subsection, the above definitions and processes related to the detection of anomalies are illustrated by Example 1.

Example 1. A mixed data table is shown in Table 1, where $OB = \{o_1, o_2, o_3, o_4, o_5\}$ and $AT = \{a_1, a_2, a_3\}$. In AT, a_1 is a nominal attribute, and a_2, a_3 are numeric attributes.

First, the raw numerical attribute data is standardized to [0, 1] by using the min–max standardization method. In order to handle mixed attribute data efficiently, for any $a \in AT$, the fuzzy similarity relations on a are calculated by the following way.

$$R_a(o,p) = \begin{cases} 1, a(o) = a(p) \text{ and } a \text{ is nominal;} \\ 0, a(o) \neq a(p) \text{ and } a \text{ is nominal;} \\ \max(\min(\frac{a(p)-a(o)+\sigma}{\sigma}, \frac{a(o)-a(p)+\sigma}{\sigma}), 0), a \text{ is numerical,} \end{cases} \tag{9}$$

where σ is the adjustable threshold.

Let $\sigma = 0.5$. According to Eq. (2), the fuzzy-rough granule structures regarding each attribute in AT are computed as follows.

$G(R_{a_1}) = \{(1, 1, 0, 0, 0), (1, 1, 0, 0, 0), (0, 0, 1, 0, 0), (0, 0, 0, 1, 1), (0, 0, 0, 1, 1)\}$;

$G(R_{a_2}) = \{(1, 0.33, 0, 0.33, 0), (0.33, 1, 0, 1, 0), (0, 0, 1, 0, 1), (0.33, 1, 0, 1, 0), (0, 0, 1, 0, 1)\}$;

$G(R_{a_3}) = \{(1, 0, 0, 0.60, 0), (0, 1, 0.60, 0, 0.60), (0, 0.60, 1, 0.20, 0.20), (0.60, 0, 0.20, 1, 0), (0, 0.60, 0.20, 0, 1)\}$.

According to Definition 4, the fuzzy-rough densities are calculated for each object with respect to a single attribute, as shown in Table 2.

By Definition 5, we have $FE(a_1) \approx 1.52$, $FE(a_2) \approx 1.29$, $FE(a_3) \approx 1.42$. Obviously, a_1 achieves the maximum fuzzy entropy value, which means that a_1 contains the maximum uncertainty and should provide more anomalous features. Therefore, it should be given more weight in the process of detecting anomalies. Further, by Eq. (8), the weights of each attribute is calculated as $W(a_1) \approx 0.36$, $W(a_2) \approx 0.30$, and $W(a_3) \approx 0.34$. Finally, by Definition 6, the anomaly scores of each object is calculated as follows.

$$AS(o_1) = W(a_k)\sum_{k=1}^{m}(1 - FRD_{a_k}(o)) = 0.36 \times (1 - 0.40) + 0.30 \times (1 - 0.33) + 0.34 \times (1 - 0.32) \approx 0.65$$

Similarly, we can get $AS(o_2) \approx 0.57$, $AS(o_3) \approx 0.67$, $AS(o_4) \approx 0.75$, $AS(o_5) \approx 0.59$, $AS(o_6) \approx 0.61$.

4.3. Detection algorithm

In the previous subsection, we construct a weighted fuzzy-rough density-based anomaly detection model. This subsection mainly designs the corresponding WFRDA algorithm.

Algorithm 1: WFRDA

Input: Data table $DT = \langle OB, AT\rangle$, σ
Output: Anomaly score AS

1 $AS \leftarrow \emptyset$;
2 **for** $a \in AT$ **do**
3 Compute $G(R_a)$ by Eq. (9);
4 Compute $FE(a)$ by Definition 5;
5 **end**
6 **for** $o \in OB$ **do**
7 **for** $a \in AT$ **do**
8 Compute FRD_a by Eq. (5);
9 Compute $W(a)$ by Eq. (8);
10 **end**
11 Compute $AS(o)$ by Definition 6;
12 **end**
13 **return** AS.

In Algorithm 1, $AS = \emptyset$ is first initialized. Second, Steps 2–5 calculate the granule structure and fuzzy entropy of each attribute $a \in AT$ through a "for" loop. Then, two "for" loops are used to calculate the fuzzy-rough density and the corresponding weights for each attribute. Finally, in Step 11, the anomaly score of each object is calculated by Definition 5 and the anomaly score AS is output in Step 13. Through analysis, the total number of rounds of Algorithm 1 is $|AT| \times |OB| \times |OB| + |OB| \times |AT|$. Therefore, in the worst case, the time complexity of Algorithm 1 is $O(|AT||OB|^2)$.

5. Experiments

This section performs an extensive experimental comparison with nine anomaly detection algorithms. Before analyzing the experimental results, we first present some preparations about the experiments, such as datasets, comparing algorithms, and evaluation indexes.

5.1. Experimental preparations

5.1.1. Datasets

To confirm the effectiveness of the proposed algorithm WFRDA, a series of comparative experiments are performed on publicly available datasets. These datasets are derived from a number of publicly available web pages,[1,2] which are frequently used in several anomaly detection research works to evaluate the detection performance. As shown in Table 3, the basic information of the experimental data is summarized. By looking at Table 3, it can be seen that the sample size of the data used is 94 to 4781 and the dimensionality of the data is 4 to 279. In addition, some medical datasets are included in Table 3, such as datasets Arrhythmia, Diabetes, Sick, etc. In the experiments, the missing values in the data are filled by the maximum frequency method. Numerical data are normalized to [0,1] by the min–max normalization method.

[1] https://github.com/BElloney/Outlier-detection
[2] http://odds.cs.stonybrook.edu

5.1.2. Comparing algorithms

Nine popular and typical detection algorithms are written to compare the algorithm WFRDA proposed in this paper. Specifically, these detection algorithms are DIStance (DIS) [10], Local Distance-based Outlier Factor (LDOF) [40], Fast Angle-Based Outlier Detection (FastABOD) [41], INFLO [18], LoOP [19], DCROD [20], Empirical Cumulative-based Outlier Detection (ECOD) [9], Isolation Forest (IForest) [42], and Weight Density-Outlier Detection (WDOD) [38], which represent different detection techniques and have relatively good performance. Among them, DIS and LDOF are distance-based algorithms; INFLO, LoOP, DCROD, and WDOD are density-based algorithms; FastABOD, ECOD, and IForest are angle-based, probability-based, and ensemble-based algorithms, respectively. Finally, the computational complexities of different outlier detection algorithms are listed in Table 4, where t, ϕ, and n denote the testing data size, the subsampling size, and the number of trees. Through Table 4, it can be seen that the computational complexity of the proposed algorithm in this paper is not the lowest, but it is within the feasible range. More importantly, the subsequent experimental comparison results verify the effectiveness of the proposed algorithm.

In our experiments, LDOF, density-based and angle-based algorithms mainly involve the parameter k, so we adjust k from 1 to 60 in steps of 1 to obtain the optimal result. The number of base estimators of IForest is set to 100. For WDOD, the Fuzzy C-Means clustering discretization method is adopted to discretize the numerical attribute values and the number of discretization intervals is 3. For WFRDA, we adjust σ from 0.1 to 2 in steps of 0.1 to obtain the best result.

5.1.3. Evaluation indexes

To comprehensively evaluate the performance of the algorithm, the Receiver Operating Curve (ROC) and the Area Under Curve (AUC) indexes are adopted in the experiments [43].

In general, most anomaly detection algorithms eventually output an Anomaly Score (AS) for each sample and a score threshold is set to determine the outliers. Given a score threshold μ, let S_d denote the current set of detected anomalies, which is determined by

$$S_d = \{o | AS(o) > \mu\}. \tag{10}$$

Let S_t denote the set of true outliers in the dataset. The True Positive Rate (TPR) and the False Positive Rate (FPR) are then defined as

$$TPR(\mu) = \frac{|S_d \bigcap S_t|}{|S_t|}, \tag{11}$$

$$FPR(\mu) = \frac{|S_d - S_t|}{|OB - S_t|}. \tag{12}$$

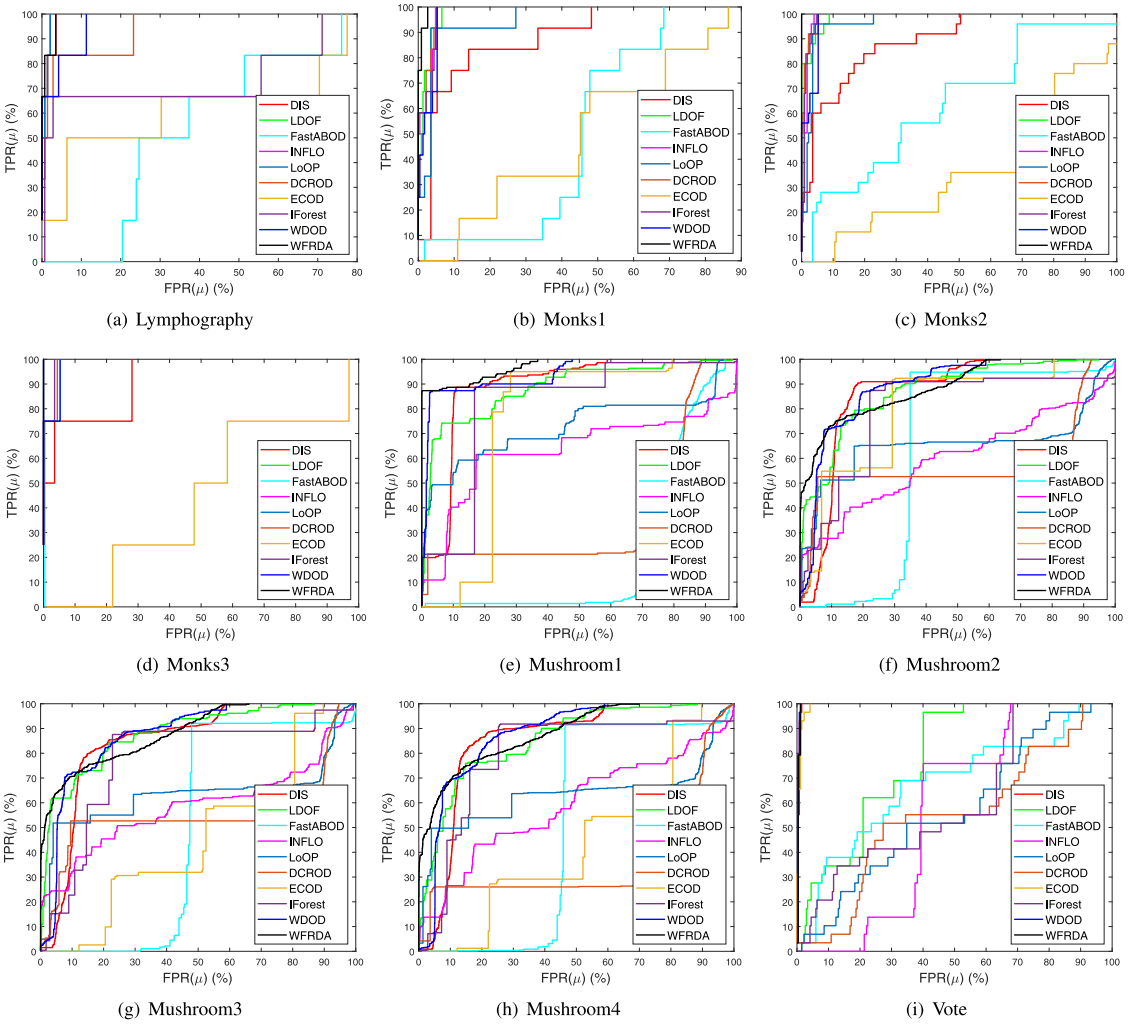
ROC is a popular evaluation metric for anomaly detection, which has the advantage of being monotonic and easier to interpret. ROC plots $FPR(\mu)$ on the x-axis and $TPR(\mu)$ on the y-axis. The closer the ROC curve of a detection algorithm is to the upper left corner of the graph, the higher its detection performance. However, it may be difficult to determine which algorithm is absolutely superior by ROC when the ROC curves of both algorithms are positioned at approximately the same level. For this reason, AUC is further proposed for evaluating the overall effectiveness of the algorithm. Usually, AUC can be defined by calculating the following average of all outlier–inlier point pairs in the dataset.

$$AUC = \underset{o_i \in S_t, o_j \in S_t^c}{Mean} \begin{cases} 1, & AS(o_i) > AS(o_j); \\ 0.5, & AS(o_i) = AS(o_j); \\ 0, & AS(o_i) < AS(o_j). \end{cases} \tag{13}$$

AUC takes values between 0 and 1. The closer AUC of an algorithm is to 1, the better its detection performance is. Furthermore, it can be seen by Formula (13) that AUC is a mean value about the probability of all sample pairs, so it does not require any additional parameter settings.

Table 3
Experimental datasets.

No.	Dataset	Abbr.	Attribute	Sample	Anomaly	Type
1	Arrhythmia_variant1	Arrhythmia	279	452	66	Mixed
2	Autos_variant1	Autos	25	205	25	Mixed
3	Cardiotocography_2and3_33_variant1	Cardio	21	1688	33	Numeric
4	Diabetes_tested_positive_26_variant1	Diabetes	8	526	26	Numeric
5	German_1_14_variant1	German	20	714	14	Mixed
6	Heart_2_16_variant1	Heart	13	166	16	Mixed
7	Hepatitis_2_9_variant1	Hepatitis	19	94	9	Mixed
8	Iris_Irisvirginica_11_variant1	Iris	4	111	11	Numeric
9	Lymphography	Lymphography	18	148	6	Nominal
10	Monks_0_12_variant1	Monks1	6	240	12	Nominal
11	Monks_0_25_variant1	Monks2	6	253	25	Nominal
12	Monks_0_4_variant1	Monks3	6	232	4	Nominal
13	Mushroom_p_221_variant1	Mushroom1	22	4429	221	Nominal
14	Mushroom_p_365_variant1	Mushroom2	22	4573	365	Nominal
15	Mushroom_p_467_variant1	Mushroom3	22	4675	467	Nominal
16	Mushroom_p_573_variant1	Mushroom4	22	4781	573	Nominal
17	Musk	Musk	166	3062	97	Numeric
18	Pageblocks_1_258_variant1	Pageblocks	10	5171	258	Numeric
19	Pima_TRUE_55_variant1	Pima	9	555	55	Numeric
20	Sick_sick_72_variant1	Sick	29	3613	72	Mixed
21	Vote_republican_29_variant1	Vote	16	296	29	Nominal
22	Wbc_malignant_39_variant1	Wbc	9	483	39	Numeric
23	Wdbc_M_39_variant1	Wdbc	31	396	39	Numeric
24	Wine	Wine	13	129	10	Numeric

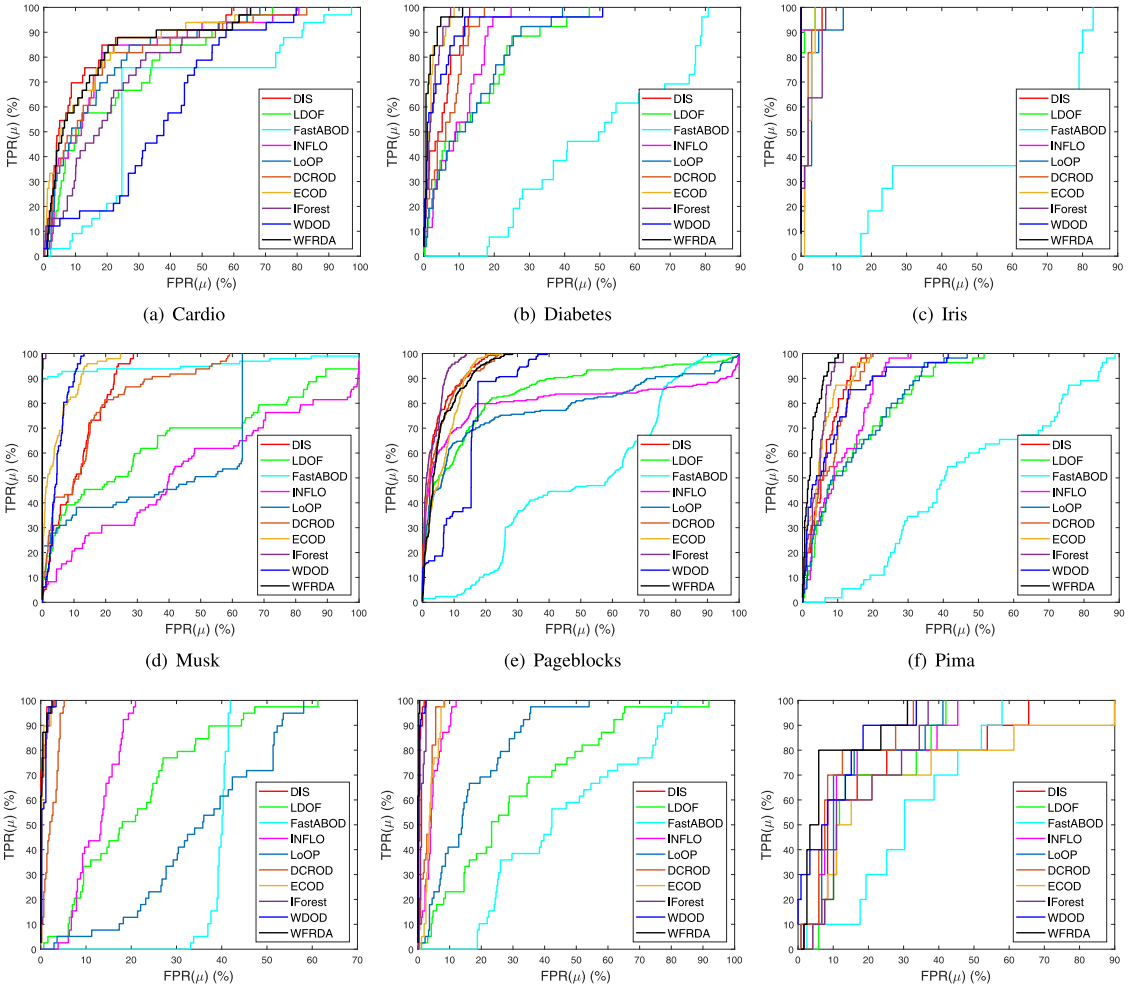


(a) Lymphography (b) Monks1 (c) Monks2

(d) Monks3 (e) Mushroom1 (f) Mushroom2

(g) Mushroom3 (h) Mushroom4 (i) Vote

Fig. 1. ROC on nominal datasets.

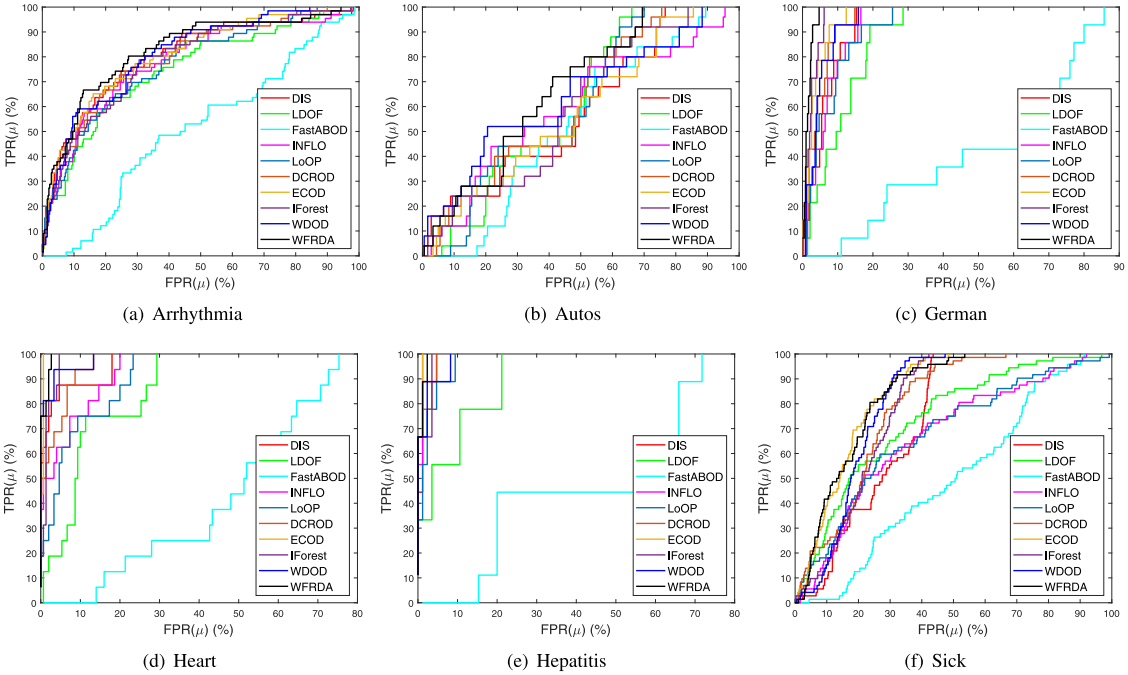


(a) Cardio (b) Diabetes (c) Iris

(d) Musk (e) Pageblocks (f) Pima

(g) Wbc (h) Wdbc (i) Wine

Fig. 2. ROC on numerical datasets.



(a) Arrhythmia (b) Autos (c) German

(d) Heart (e) Hepatitis (f) Sick

Fig. 3. ROC on nominal datasets.

Table 4
The computational complexities of different algorithms.

Algorithm	Time complexity
DIS	$O(\|AT\|\|OB\|^2)$
LDOF	$O(\|AT\|\|OB\|^2)$
FastABOD	$O(\|AT\|(\|OB\|^2 + t^2\|OB\|))$
INFLO	$O(\|AT\|\|OB\| \log \|OB\|)$
LoOP	$O(\|AT\|\|OB\|^2)$
DCROD	$O(\|AT\|\|OB\| \log \|OB\|)$
ECOD	$O(\|AT\|\|OB\|)$
IForest	$O(t\phi^2 + nt\phi)$
WDOD	$O(\|AT\|\|OB\| \log \|OB\|)$
WFRDA	$O(\|AT\|\|OB\|^2)$

5.2. Experimental results on ROC

First, we analyze the comparison results regarding ROC. The ROC curves of ten outlier detection algorithms regarding nominal, numerical, and mixed attribute data are shown in Figs. 1–3, respectively. Among them, the black curves show the algorithms proposed in this paper. With Figs. 1–3, it is clear that the ROC curves of each comparison algorithm are monotonically undiminished. As mentioned earlier, the closer the ROC curve of a detection algorithm is to the upper left corner of the figure, the higher its detection performance is.

Through Figs. 1–3, we can see that the ROC curves of the proposed WFRDA on some datasets are clearly closest to the upper left corner of the first quadrant, such as datasets Monks2, Mushroom, Iris, Musk, Pima, German, etc. Herein, the ROC curves of WFRDA on datasets Musk and Monks2 overlap exactly with the coordinates, which indicates that its exhibits full performance on datasets Musk and Monks2. In addition, the ROC curve of WFRDA is closer to the upper left corner of the first quadrant on some other datasets, such as datasets Vote, Diabetes, Sick, etc. Therefore, we can conclude that WFRDA exhibits superior performance in most cases.

However, for some datasets, such as datasets Cardio and Wbc, the ROC curves of a certain two algorithms behave similarly. In this case, it is difficult to say which algorithm is absolutely superior. Therefore, we further give the comparison results of ten detection algorithms regarding AUC below.

5.3. Experimental results on AUC

Second, we analyze the AUC values of the comparison algorithms. The experimental comparison results on AUC are summarized in Table 5, where the bolded part indicates the best results.

Based on these twenty-four public datasets, Table 5 can better reflect the advantages of the proposed detection algorithm. Obviously, WFRDA achieves better AUC values in most datasets. For instance, the AUC value of WFRDA on dataset Cred is 0.826. However, the AUC values of DIS, LDOF, FastABOD, INFLO, LoOP, DCROD, ECOD, IForest, and WDOD are only 0.803, 0.749, 0.492, 0.782, 0.771, 0.796, 0.807, 0.788, and 0.811, respectively, which are all smaller than that of WFRDA.

From a statistical point of view, WFRDA achieves the best results on 17 datasets; the remaining algorithms DIS, LDOF, FastABOD, INFLO, LoOP, DCROD, ECOD, IForest, and WDOD achieve the best results only on 3, 1, 0, 2, 1, 0, 3, 2, and 3, respectively.

From a theoretical point of view, both substitution and discretization of data values may lead to changes in the data structure, which in turn leads to information loss. On the dataset with numerical attributes, the AUC value of WDOD is in most cases smaller than that of WFRDA proposed in this paper. The reason for this may be the discretization of the numerical attribute values. For datasets containing nominal attributes, replacing nominal attributes with different integer values also affects the performance of algorithms such as DIS, LDOF, and FastABOD. In contrast, Formula (9) can directly handle the mixed data without any transformation for WFRDA, which can ultimately retain more real information of the data. At the same time, the proposed method also effectively utilizes the advantage that the fuzzy-rough computing theory can effectively handle fuzzy information. Therefore, WFRDA achieves the best AUC values on most data sets.

That is to say, WFRDA is able to achieve better results not only in experimental results but also its superiority is confirmed by theoretical analysis. The final average of the AUC values also can better reveal and validate comparative performance. DIS, LDOF, FastABOD, INFLO, LoOP, DCROD, ECOD, IForest, WDOD, and WFRDA correspond to 0.898, 0.848, 0.592, 0.813, 0.811, 0.856, 0.794, 0.881, 0.912, and 0.941, respectively. Among them, WFRDA obtains the best value of 0.941, which is significantly larger than those of other algorithms.

In addition, these datasets include nominal data, numerical data, and mixed attribute data. In conclusion, through comparative analyses, WFRDA can effectively find outliers in nominal, numerical, and mixed data.

5.4. Effect of attribute noise on AUC

Third, we use the strategy in [44] to evaluate the effect of attribute noise on AUC. Specifically, error values are introduced into each attribute through a level $x \cdot 100\%$. To obtain a data set with noise level $x \cdot 100\%$, $\lceil xn \rceil$ samples at a certain attribute value are replaced with a random value, respectively. In this case, for the numerical attribute, a random value between the maximum and minimum values is selected to replace the original value. For nominal attributes, a different nominal attribute value is randomly selected to replace the original value.

The final results of the effect of attribute noise level on AUC are shown in Fig. 4. From the experimental results in Fig. 4, we can see that as the noise level increases, the AUC tends to fluctuate up and down on most of the datasets, but their fluctuations are not too much, such as datasets Wdbc, Wbc, Vote, German, etc. This indicates that the proposed WFRDA has some robustness to attribute noise.

5.5. Statistical analysis

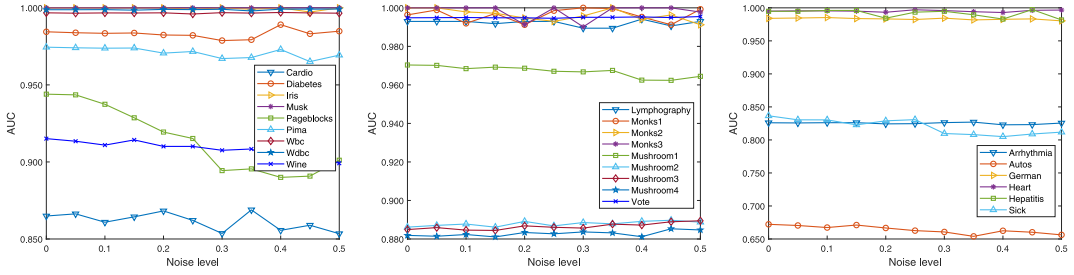
Fourth, Friedman's test [45] and Nemenyi's post-hoc test [46] are applied to evaluate the statistical significance of the results. Before using Friedman's test, AUC value of each algorithm on all datasets is sorted from low to high, and the sequence number is assigned $(1, 2, \ldots)$. Among them, if AUC value of the two algorithms is the same, the ordinal values are equally divided. Then, Friedman's test is used to determine whether these algorithms have the same performance. Suppose we compare M algorithms on N datasets, and let r_i denote the average ordinal value of the ith algorithm, then Friedman's test is calculated as follows.

$$\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(M-1)-\tau_{\chi^2}} \text{ and}$$
$$\tau_{\chi^2} = \frac{12N}{M(M+1)}\left(\sum_{i=1}^{M} r_i^2 - \frac{M(M+1)^2}{4}\right). \quad (14)$$

τ_F obeys the F distribution with $(M-1)$ and $(M-1)(N-1)$ degrees of freedom. If the null hypothesis of "all algorithms have the same performance" is rejected, it means that the performance

Table 5
Experimental comparison results on AUC.

Datasets	DIS	LDOF	FastABOD	INFLO	LoOP	DCROD	ECOD	IForest	WDOD	WFRDA
Arrhythmia	0.803	0.749	0.492	0.782	0.771	0.796	0.807	0.788	0.811	**0.826**
Autos	0.587	0.616	0.536	0.600	0.610	0.619	0.583	0.596	0.641	**0.672**
Cardio	**0.877**	0.790	0.519	0.837	0.833	0.834	0.871	0.788	0.641	0.865
Diabetes	0.952	0.863	0.450	0.902	0.870	0.935	0.979	0.976	0.951	**0.984**
German	0.938	0.889	0.496	0.932	0.925	0.955	0.966	0.975	0.953	**0.984**
Heart	0.970	0.880	0.548	0.945	0.923	0.970	**0.996**	0.984	0.987	0.995
Hepatitis	0.990	0.922	0.653	0.995	0.976	0.990	**0.996**	0.988	0.986	0.995
Iris	**1.000**	0.995	0.541	0.995	0.969	0.983	0.977	0.971	**1.000**	**1.000**
Lymphography	0.989	0.992	0.596	**0.994**	0.993	0.954	0.682	0.781	0.974	0.993
Monks1	0.891	0.984	0.649	0.982	0.955	0.986	0.539	0.984	0.976	**0.996**
Monks2	0.892	0.989	0.660	0.987	0.970	0.992	0.374	0.988	0.980	**1.000**
Monks3	0.921	**1.000**	0.911	**1.000**	**1.000**	0.989	0.440	0.992	0.987	**1.000**
Mushroom1	0.893	0.887	0.502	0.632	0.753	0.596	0.759	0.824	0.941	**0.970**
Mushroom2	0.866	0.873	0.646	0.598	0.753	0.728	0.816	0.812	**0.890**	0.886
Mushroom3	0.847	0.882	0.515	0.587	0.718	0.710	0.465	0.784	0.876	**0.885**
Mushroom4	0.851	0.855	0.524	0.590	0.718	0.604	0.437	0.805	0.875	**0.882**
Musk	0.889	0.663	0.954	0.534	0.557	0.863	0.956	**1.000**	0.950	**1.000**
Pageblocks	0.956	0.849	0.513	0.807	0.791	0.953	0.938	**0.970**	0.866	0.944
Pima	0.937	0.855	0.492	0.894	0.860	0.928	0.947	0.957	0.919	**0.974**
Sick	0.731	0.739	0.530	0.673	0.681	0.788	**0.843**	0.784	0.808	0.837
Vote	0.995	0.780	0.680	0.568	0.500	0.537	0.993	0.603	**0.995**	**0.995**
Wbc	**0.997**	0.792	0.604	0.875	0.645	0.976	0.995	0.996	0.993	**0.997**
Wdbc	0.995	0.702	0.523	0.954	0.839	0.968	0.959	0.987	0.997	**0.999**
Wine	0.797	0.816	0.681	0.850	0.855	0.885	0.733	0.824	0.900	**0.915**
Average	0.898	0.848	0.592	0.813	0.811	0.856	0.794	0.881	0.912	**0.941**



(a) Numeric (b) Nominal (c) Mixed

Fig. 4. Effect of attribute noise on AUC.

of the algorithms is significantly different. At this time, a post-hoc test needs to be used to further distinguish these feature selection algorithms. Nemenyi's post-hoc test is commonly used. In Nemenyi's test, the critical difference (CD) of the average ordinal value is calculated by the following formula.

$$CD_\alpha = q_\alpha \sqrt{\frac{M(M+1)}{6N}}, \tag{15}$$

where q_α is the critical value of Tukey's distribution, which can be found in [46].

Further, Nemenyi's test figure is used to more intuitively represent the significant differences between the two algorithms [27]. In Nemenyi's test figure, for each algorithm, a dot is used to show its average ordinal value, and a horizontal line segment with the dot as the center is used to indicate the size of CD. If a group of algorithms is connected by horizontal line segments, then it means that there is no significant difference between this group of algorithms.

Specifically, we can get $M = 10$ and $N = 24$, the τ_F distribution has 9 and 207 degrees of freedom. According to Friedman's test, when $\alpha = 0.05$, the value of τ_F is greater than the critical value 1.9253. Therefore, the null hypothesis that "all algorithms have the same performance" is rejected. It shows that the performance of all outlier detection algorithms is significantly different. At this time, a post-hoc test needs to be used to further distinguish them.

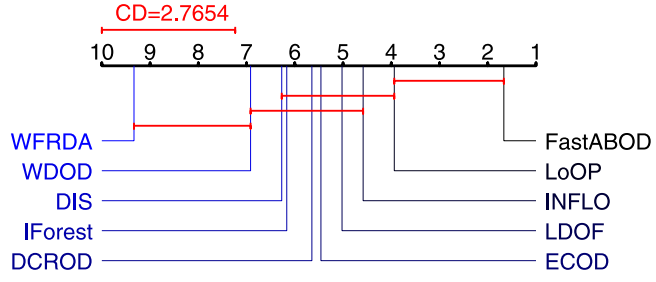
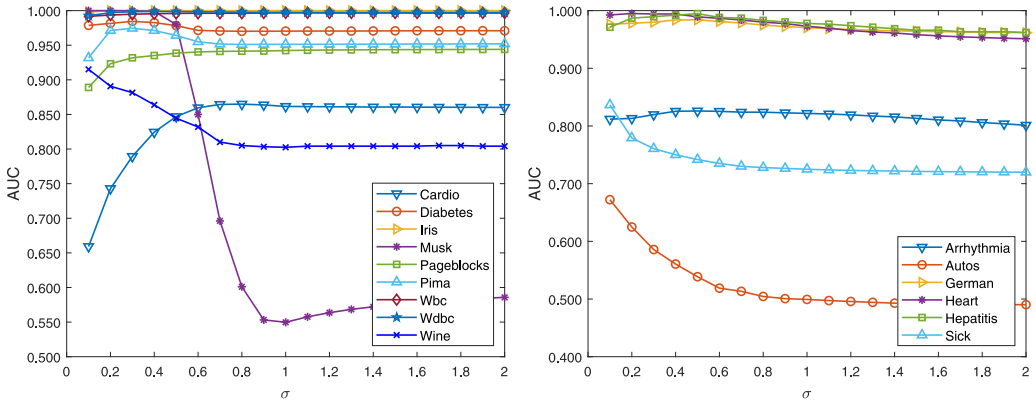


Fig. 5. Nemenyi's test figures on AUC.

For significance level $\alpha = 0.05$, the corresponding critical distance $CD_{0.05} = 2.7654$ can be obtained. Finally, Nemenyi's test figure on AUC is shown in Fig. 5. From Fig. 5, we can see that WFRDA is statistically significantly different from most other algorithms. For example, it can be seen from Fig. 5 that WFRDA is not connected to DIS, LDOF, FastABOD, INFLO, LoOP, DCROD, ECOD, and IForest with horizontal line segments, which indicates that WFRDA is statistically significantly different from these algorithms. However, there is no consistent evidence to indicate the statistical differences from WDOD.

5.6. Parameter sensitivity analyses

Finally, we explore the sensitivity of WFRDA to the parameters σ. The plots of AUC with respect to parameters σ for numerical



(a) Numeric (b) Mixed

Fig. 6. Variation curve of AUC with parameter σ.

and mixed data are drawn in Figs. 6(a) and 6(b), respectively. Through Figs. 6(a) and 6(b), we can see that the AUC values first increase with increasing σ on some datasets and then gradually level off, such as the datasets Diabetes, Iris, Pima, etc. This indicates that WFRDA is less sensitive to the parameter σ after it reaches a certain value. In addition, on some datasets, such as datasets Musk and Autos, the AUC values first level off and then gradually decreas as σ increases. This indicates that WFRDA is sensitive on datasets Musk and Autos. However, in either case, the proposed WFRDA achieves better AUC values on the appropriate parameter values.

6. Conclusions

To address the shortcomings of existing density-based methods, this paper extends the idea of weight density to FRS theory, and then defines anomalies based on weight fuzzy-rough density. In the concept of anomalies proposed in this paper, the fuzzy relations are calculated by a mixed measurement. In addition, this paper fully considers the advantage of the fuzzy-rough computing model that can effectively handle fuzzy data to calculate anomaly scores. As a result, the proposed detection model is not only effective in handling fuzzy information but also applicable to mixed data. The comparative analysis of experimental results verifies that WFRDA can effectively and comprehensively detect outliers. In future work, we will further explore the fuzzy-rough computing method for collective anomaly detection.

CRediT authorship contribution statement

Zhong Yuan: Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Baiyang Chen:** Formal analysis, Data curation, Software. **Jia Liu:** Formal analysis, Data curation, Software. **Hongmei Chen:** Project administration, Validation, Funding acquisition. **Dezhong Peng:** Project administration, Validation, Funding acquisition. **Peilin Li:** Project administration, Validation, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets and codes are publicly available online at https://github.com/BElloney/Outlier-detection.

Acknowledgments

The authors thank both the editors and reviewers for their valuable suggestions, which substantially improve this paper. This work was supported by the National Natural Science Foundation of China (U19A2078 and 61976182), Sichuan Science and Technology Planning Project (2023YFQ0020, 2022YFQ0014, 2022YFH0021, 2022YFS0128, 2021YFS0389, and 2021YFS0390), and the Fundamental Research Funds for the Central Universities, China (YJ202245).

References

[1] H.F. Liu, J. Li, Y. Wu, Y. Fu, Clustering with outlier removal, IEEE Trans. Knowl. Data Eng. 33 (6) (2019) 2369–2379.
[2] C. Chen, Y. Wang, W.B. Hu, Z.B. Zheng, Robust multi-view k-means clustering with outlier removal, Knowl.-Based Syst. 210 (2020) 106518.
[3] B. Wang, Z.Z. Mao, Outlier detection based on Gaussian process with application to industrial processes, Appl. Soft Comput. 76 (2019) 505–516.
[4] S.Y. Liu, Y.X. Zhao, Z.Z. Lin, Y.L. Liu, Y. Ding, L. Yang, S.M. Yi, Data-driven event detection of power systems based on unequal-interval reduction of PMU data and local outlier factor, IEEE Trans. Smart Grid 11 (2) (2019) 1630–1643.
[5] J. Liu, T.R. Li, Z. Yuan, W. Huang, P. Xie, Q. Huang, Symbolic aggregate approximation based data fusion model for dangerous driving behavior detection, Inform. Sci. 609 (2022) 626–643.
[6] T. Pourhabibi, K.-L. Ong, B.H. Kam, Y.L. Boo, Fraud detection: A systematic literature review of graph-based anomaly detection approaches, Decis. Support Syst. 133 (2020) 113303.
[7] T. Fernando, S. Denman, D. Ahmedt-Aristizabal, S. Sridharan, K.R. Laurens, P. Johnston, C. Fookes, Neural memory plasticity for medical anomaly detection, Neural Netw. 127 (2020) 67–81.
[8] M. Goldstein, A. Dengel, Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm, in: KI-2012: Poster and Demo Track, 9, Citeseer, 2012.
[9] Z. Li, Y. Zhao, X.Y. Hu, N. Botta, C. Ionescu, G. Chen, ECOD: Unsupervised outlier detection using empirical cumulative distribution functions, IEEE Trans. Knowl. Data Eng. http://dx.doi.org/10.1109/TKDE.2022.3159580.
[10] E.M. Knox, R.T. Ng, Algorithms for mining distancebased outliers in large datasets, in: Proceedings of the International Conference on Very Large Data Bases, Citeseer, 1998, pp. 392–403.
[11] M. Radovanović, A. Nanopoulos, M. Ivanović, Reverse nearest neighbors in unsupervised distance-based outlier detection, IEEE Trans. Knowl. Data Eng. 27 (5) (2014) 1369–1382.
[12] M.M. Breunig, H.-P. Kriegel, R.T. Ng, J. Sander, LOF: identifying density-based local outliers, in: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, pp. 93–104.
[13] B. Tang, H.B. He, A local density-based approach for outlier detection, Neurocomputing 241 (2017) 171–180.
[14] Z.Y. He, X.F. Xu, S.C. Deng, Discovering cluster-based local outliers, Pattern Recognit. Lett. 24 (9–10) (2003) 1641–1650.
[15] J.B. Li, H. Izakian, W. Pedrycz, I. Jamal, Clustering-based anomaly detection in multivariate time series data, Appl. Soft Comput. 100 (2021) 106919.
[16] J. Tang, Z.X. Chen, A.W.-C. Fu, D.W. Cheung, Enhancing effectiveness of outlier detections for low density patterns, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2002, pp. 535–548.

[17] S. Papadimitriou, H. Kitagawa, P.B. Gibbons, C. Faloutsos, LOCI: Fast outlier detection using the local correlation integral, in: Proceedings 19th International Conference on Data Engineering, IEEE, 2003, pp. 315–326.
[18] W. Jin, A.K. Tung, J.W. Han, W. Wang, Ranking outliers using symmetric neighborhood relationship, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2006, pp. 577–593.
[19] H.-P. Kriegel, P. Kröger, E. Schubert, A. Zimek, Loop: local outlier probabilities, in: Proceedings of the 18th ACM Conference on Information and Knowledge Management, 2009, pp. 1649–1652.
[20] K.S. Li, X. Gao, S.Y. Fu, X.P. Diao, P. Ye, B. Xue, J.H. Yu, Z.J. Huang, Robust outlier detection based on the changing rate of directed density ratio, Expert Syst. Appl. 207 (2022) 117988.
[21] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[22] Q.H. Hu, D.R. Yu, W. Pedrycz, D.G. Chen, Kernelized fuzzy rough sets and their applications, IEEE Trans. Knowl. Data Eng. 23 (11) (2010) 1649–1667.
[23] C.Z. Wang, M.W. Shao, Q. He, Y.H. Qian, Y.L. Qi, Feature subset selection based on fuzzy neighborhood rough sets, Knowl.-Based Syst. 111 (2016) 173–179.
[24] J.M. Zhan, H.B. Jiang, Y.Y. Yao, Covering-based variable precision fuzzy rough sets with PROMETHEE-EDAS methods, Inform. Sci. 538 (2020) 314–336.
[25] S. An, Q.H. Hu, C.Z. Wang, Probability granular distance-based fuzzy rough set model, Appl. Soft Comput. 102 (2021) 107064.
[26] C.Z. Wang, Y. Huang, M.W. Shao, X.D. Fan, Fuzzy rough set-based attribute reduction using distance measures, Knowl.-Based Syst. 164 (2019) 205–212.
[27] Z. Yuan, H.M. Chen, P. Xie, P.F. Zhang, J. Liu, T.R. Li, Attribute reduction methods in fuzzy rough set theory: An overview, comparative experiments, and new directions, Appl. Soft Comput. 107 (2021) 107353.
[28] J.M. Zhan, B.Z. Sun, J.C.R. Alcantud, Covering based multigranulation (I,T)-fuzzy rough set models and applications in multi-attribute group decision-making, Inform. Sci. 476 (2019) 290–318.
[29] W.J. Wang, J.M. Zhan, C. Zhang, E. Herrera-Viedma, G. Kou, A regret-theory-based three-way decision method with a priori probability tolerance dominance relation in fuzzy incomplete information systems, Inf. Fusion 89 (2023) 382–396.
[30] J.J. Wang, X.L. Ma, Z.S. Xu, J.M. Zhan, Regret theory-based three-way decision model in hesitant fuzzy environments and its application to medical decision, IEEE Trans. Fuzzy Syst. 30 (12) (2022) 5361–5375.
[31] J.J. Wang, X.L. Ma, Z.S. Xu, W. Pedrycz, J.M. Zhan, A three-way decision method with prospect theory to multi-attribute decision-making and its applications under hesitant fuzzy environments, Appl. Soft Comput. 126 (2022) 109283.
[32] J.M. Zhan, J.J. Wang, W.P. Ding, Y.Y. Yao, Three-Way Behavioral Decision Making With Hesitant Fuzzy Information Systems: Survey and Challenges, IEEE/CAA J. Autom. Sin. http://dx.doi.org/10.1109/JAS.2022.106061.
[33] S.L. Xu, S.L. Liu, J. Zhou, L. Feng, Fuzzy rough clustering for categorical data, Int. J. Mach. Learn. Cybern. 10 (11) (2019) 3213–3223.
[34] S.Y. Zhao, Z.G. Dai, X.Z. Wang, P. Ni, H.H. Luo, H. Chen, C.P. Li, An accelerator for rule induction in fuzzy rough theory, IEEE Trans. Fuzzy Syst. 29 (12) (2021) 3635–3649.
[35] Z. Yuan, H.M. Chen, T.R. Li, J. Liu, S. Wang, Fuzzy information entropy-based adaptive approach for hybrid feature outlier detection, Fuzzy Sets and Systems 421 (2021) 1–28.
[36] Z. Yuan, H.M. Chen, T.R. Li, B.B. Sang, S. Wang, Outlier detection based on fuzzy rough granules in mixed attribute data, IEEE Trans. Cybern. 52 (8) (2022) 8399–8412.
[37] F.Y. Cao, J.Y. Liang, L. Bai, A new initialization method for categorical data clustering, Expert Syst. Appl. 36 (7) (2009) 10223–10228.
[38] X.W. Zhao, J.Y. Liang, F.Y. Cao, A simple and effective outlier detection algorithm for categorical data, Int. J. Mach. Learn. Cybern. 5 (3) (2014) 469–477.
[39] Q.H. Hu, D.R. Yu, Z.X. Xie, J.F. Liu, Fuzzy probabilistic approximation spaces and their information measures, IEEE Trans. Fuzzy Syst. 14 (2) (2006) 191–201.
[40] K. Zhang, M. Hutter, H. Jin, A new local distance-based outlier detection approach for scattered real-world data, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2009, pp. 813–822.
[41] H.-P. Kriegel, M. Schubert, A. Zimek, Angle-based outlier detection in high-dimensional data, in: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008, pp. 444–452.
[42] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation forest, in: 2008 Eighth Ieee International Conference on Data Mining, IEEE, 2008, pp. 413–422.
[43] G.O. Campos, A. Zimek, J. Sander, R.J. Campello, B. Micenková, E. Schubert, I. Assent, M.E. Houle, On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study, Data Min. Knowl. Discov. 30 (4) (2016) 891–927.
[44] X.Q. Zhu, X.D. Wu, Class noise vs. attribute noise: A quantitative study, Artif. Intell. Rev. 22 (3) (2004) 177–210.
[45] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1) (1940) 86–92.
[46] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (Jan) (2006) 1–30.