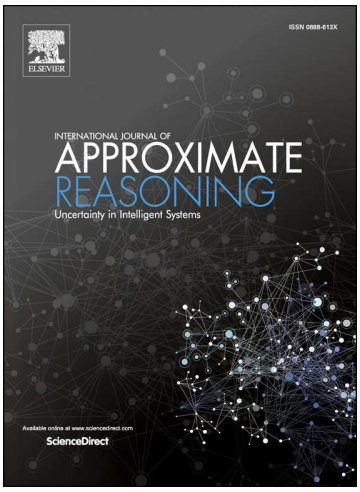
Journal Pre-proof



Exploiting fuzzy rough entropy to detect anomalies

Sihan Wang, Zhong Yuan, Chuan Luo, Hongmei Chen and Dezhong Peng

PII: S0888-613X(23)00218-9
DOI: https://doi.org/10.1016/j.ijar.2023.109087
Reference: IJA 109087

To appear in: *International Journal of Approximate Reasoning*

Received date: 27 August 2023
Revised date: 22 October 2023
Accepted date: 14 November 2023

Please cite this article as: S. Wang, Z. Yuan, C. Luo et al., Exploiting fuzzy rough entropy to detect anomalies, *International Journal of Approximate Reasoning*, 109087, doi: https://doi.org/10.1016/j.ijar.2023.109087.

Exploiting fuzzy rough entropy to detect anomalies

Sihan Wang[a], Zhong Yuan[a,*], Chuan Luo[a], Hongmei Chen[b], Dezhong Peng[a]

[a]College of Computer Science, Sichuan University, Chengdu 610065, China
[b]School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China

Abstract

Anomaly detection has been used in a wide range of fields. However, most of the current detection methods are only applicable to certain data, ignoring uncertain information such as fuzziness in the data. Fuzzy rough set theory, as an essential mathematical model for granular computing, provides an effective method for processing uncertain data such as fuzziness. Fuzzy rough entropy has been proposed in fuzzy rough set theory and has been employed successfully in data analysis tasks such as feature selection. However, it mainly uses the intersection operation, which may not effectively reflect the similarity between high-dimensional objects. In response to the two challenges mentioned above, distance-based fuzzy rough entropy and its correlation measures are proposed. Further, the proposed fuzzy rough entropy is used to construct the anomaly detection model and the Fuzzy Rough Entropy-based Anomaly Detection (FREAD) algorithm is designed. Finally, the FREAD algorithm is compared and analyzed with some mainstream anomaly detection algorithms (including COF, DIS, INFLO, LDOF, LoOP, MIX, ODIN, SRO, and VarE algorithms) on some publicly available datasets. Experimental results indicate that the FREAD algorithm significantly outperforms other algorithms in terms of performance and flexibility. The code is publicly available online at https://github.com/optimusprimeyy/FREAD.

Keywords: Anomaly detection; Outlier detection; Fuzzy rough entropy; Uncertainty measure; Mixed data

1. Introduction

Anomalies (also known as outliers) are the few sample points in the data set that are significantly different from the majority of other samples. Currently, anomaly detection is crucial in many areas, including electricity anomaly detection [1], credit card fraud detection [2], software defect prediction [3], and intrusion detection [4]. An anomaly is not equivalent to an incorrect data object, which is usually given a special meaning. For instance, it might signal significant occurrences like credit card fraud, outside network infiltration, and user energy stealing. Therefore, it is crucial to research more effective anomaly detection models.

Existing anomaly detection methods are mainly divided into four categories: statistical-based [5, 6], cluster-based [7, 8], proximity-based [9, 10] (including distance-based and density-based) and rough set-based methods [11]. Classical anomaly detection methods are mainly suitable for deterministic numerical data, but there are plenty of uncertain and incomplete data in real life. These data may also store a lot of valuable and mineable information. In order to deal with this kind of data, scholars have proposed a series of rough set-based anomaly detection methods [11–13]. However, classical rough set-based methods are based on equivalence relations and equivalence classes. So their detection models are only available for nominal data, but not directly to numerical data. When these methods are applied to datasets containing numerical attributes, the numerical attribute data must be discretized. This results in a substantial addition in data processing time and information loss during discretization, which affects the detection performance of the model. In view of the shortcomings of the classical rough set model, scholars have attempted to detect outliers using some extended rough set models. For example, the neighborhood rough set model was introduced to detect anomalies [14, 15]. Neighborhood rough sets can directly handle numerical data, thus the information loss induced by the discretization process is avoided. However, in practical applications, there is usually fuzzy information in the data. The existing anomaly detection methods based on the neighborhood rough set model need to improve the detection performance because they do not consider the fuzzy information of the data. Fuzzy rough set theory has

*Corresponding author
Email addresses: wangsihan0713@foxmail.com (Sihan Wang), yuanzhong@scu.edu.cn (Zhong Yuan), cluo@scu.edu.cn (Chuan Luo), hmchen@swjtu.edu.cn (Hongmei Chen), pengdz@scu.edu.cn (Dezhong Peng)

been extensively employed in tasks of data mining [16–22], and it is a granular computing model that can effectively handle fuzzy information. Fuzzy rough entropy, as an essential uncertainty measure in fuzzy rough set theory, has been used for feature selection [17]. The feature selection based on fuzzy rough entropy can obviate the issue of information loss brought on by discretization and effectively addresses the defects of feature selection in the rough set model, thus improving the performance of algorithms such as classification.

However, the current fuzzy rough entropy mainly uses the intersection operation, which may not accurately reflect the similarity between samples in the high-dimensional space [23]. In the field of anomaly detection, the fuzzy rough entropy may not be able to accurately detect anomalies in datasets since it cannot effectively distinguish the differences of the samples. Hence, the existing fuzzy rough entropy still needs further improvement in the construction strategy. For this purpose, this paper first proposes distance-based fuzzy rough entropy and its related measures (joint entropy, conditional entropy, mutual information entropy, etc.), and further uses the proposed fuzzy rough entropy to detect outliers and develop a related Fuzzy Rough Entropy-based Anomaly Detection (FREAD) algorithm. The proposed fuzzy rough entropy can effectively measure the similarity between samples. The proposed method can fully utilize the advantages of fuzzy rough sets, which suits nominal, numerical, and mixed attribute datasets. To prove its effectiveness, experiments are in comparison with some main outlier detection methods on publicly available datasets. Finally, our main contributions in this paper are as follows.

1) A hybrid distance metric is defined to construct the fuzzy granular structure.
2) Distance-based fuzzy rough entropy and its related measures are proposed.
3) Two kinds of attribute sequences are constructed to define the anomaly scores.
4) An anomaly detection model based on fuzzy rough entropy is proposed.
5) Experimental results show that the FREAD algorithm is significantly better than other existing anomaly detection algorithms in terms of performance and flexibility.

The structure of this paper is organized as follows. Section 2 will briefly introduce the current research status of anomaly detection methods. Section 3 then reviews some knowledge of the fuzzy rough set involved in this paper. Section 4 will propose a new distance-based fuzzy rough entropy and its related measures. In Section 5, we proceed to the presentation of our proposed anomaly detection method. The performance evaluation of different anomaly detection methods is shown in Section 6. Finally, the conclusion of our paper is provided.

2. Related work

Among the classical anomaly detection methods, the statistical-based approach was the first to be suggested and employed in the field of statistics. According to its idea [6, 24], objects are classified as outliers if they do not fit the model or diverge significantly from the statistical model in terms of the shape of the distribution. Therefore, it is highly dependent on the distribution and type of data, and is not applicable to data sets with unknown distribution. Knorr et al. [9, 25] first proposed the distance-based outlier detection method. An object can be judged as an outlier if most of the other objects in the dataset do not lie within its neighborhood. Nevertheless, it is difficult to detect local outliers in the dataset. In response to this issue, Breunig et al. [10] presented a density-based outlier detection method. The method defines the concept of local density based on factors such as relative distance and the number of objects within k-nearest neighbors, thus defining a local outlier factor for each object. The larger the outlier factor of a data object, the higher the probability that the data object will be judged as an outlier. Some researchers also proposed a cluster-based outlier detection method [7]. The method first divides all objects in the dataset into clusters by clustering. The data objects will be considered as outliers if they are located far from their clusters. However, the detection models are usually constructed using the Euclidean metric in density-based and cluster-based methods. Therefore, they are often not the best choices when detecting outliers in nominal or mixed attribute data.

Advances in technology have led to changes in data types and thus to datasets dominated by nominal or mixed attributes, which makes many traditional methods no longer applicable to this situation. As a result, many researchers have studied outlier detection methods based on rough sets [26, 27]. A new definition of outliers in rough set theory, i.e., sequence-based outliers, was proposed in [11]. Suri et al. [28] proposed a clustering algorithm for detecting nominal data by modifying the traditional k-modes algorithm, which addresses the uncertainty in the clustering process by considering a soft computational approach based on rough sets. For the issue that the proximity-based method makes it difficult to determine the size of the nearest neighbor, Jiang et al. [29] presented an outlier detection method based on the entropy of approximate accuracy. Based on an extension of the mathematical framework and rough set theory, Maciá-Pérez et al. presented an efficient algorithm for detecting outliers in extensive information in [30], which enables virtually linear complexity for the task of outlier detection.

However, since the above methods are based on equivalence relations in classical rough sets, these detection models are only suitable for nominal data. They require discretization of the data when dealing with numerical and mixed data. For addressing this problem, Chen et al. [14] proposed an outlier detection model based on neighborhood rough set theory, which combines its granularity capabilities for uncertain data with outlier identification and finds outliers in the chosen subspace. Yuan et al. [15] used heterogeneous distance and adaptive radius to determine the neighborhood information system, which defined the neighborhood information entropy to achieve the overall uncertainty measure. Wang et al. [31] further constructed a weighted neighborhood information network to represent the mixed-value attribute dataset by considering the neighborhood relations and similarity between objects. In addition, Yuan et al. [32] put up a novel outlier detection method based on multigranulation relative entropy in the neighborhood rough set, which can be applied to mixed attribute datasets by utilizing a mixed distance metric.

The current neighborhood rough set model will probably lose several important pieces of information while detecting outliers, as it does not take into account the fuzzy information of the data. To address this issue, outlier detection methods based on the fuzzy rough set have been successively proposed. These methods can effectively handle the fuzzy information within the data. Yuan et al. [33] defined the granular outlier degree (GOD) to describe the outlier degree of fuzzy rough particles by utilizing the fuzzy approximation accuracy. Furthermore, Yuan et al. [34] introduced the definition of fuzzy-rough density to describe the degree of aggregation of objects, and the anomaly detection model they developed can effectively and comprehensively detect anomalies. However, all of the above methods use intersection operations, which can result in loss of detection performance. Therefore, this paper uses distance metrics to construct fuzzy rough entropy to replace the original intersection operation, and proposes a fuzzy rough entropy-based anomaly detection algorithm to detect anomalies.

3. Preliminaries

A data matrix without decision information can be brought into a fuzzy information system $FIS = (U, C, V, f)$, where $U = \{u_1, u_2, \ldots, u_n\}$ is a non-empty finite set of objects, C is a non-empty finite set of conditional attributes. And V is the union of all attribute value domains, i.e., $V = \bigcup_{c \in C} V_c$, where V_c is the attribute c of the value domain. For any $u \in U$ and $c \in C$, $f_c(u) \in V_c$ denotes the value of the object u under the attribute c.

Definition 1 *Given a subset of attributes $B \subseteq C$, the fuzzy relation $\widetilde{R}_B$ on B over U is defined as*

$$\widetilde{R}_B : U \times U \to [0, 1]. \tag{1}$$

For any $u_i, u_j \in U$, if $\widetilde{R}_B$ satisfies the following properties:

1) Reflexivity: $\widetilde{R}_B(u_i, u_i) = 1$;
2) Symmetry: $\widetilde{R}_B(u_i, u_j) = \widetilde{R}_B(u_j, u_i)$,

then $\widetilde{R}_B$ is said to be a fuzzy similarity relation, and $\widetilde{R}_B(u_i, u_j)$ describes the similarity between samples u_i and u_j on the subset of attributes B. In particular, $\widetilde{R}_B$ is a fuzzy equivalence relation if $\widetilde{R}_B$ also satisfies transitivity, that is, for any $u_i, u_j, u_k \in U, \widetilde{R}_B(u_i, u_k) \geq \min\{\widetilde{R}_B(u_i, u_j), \widetilde{R}_B(u_j, u_k)\}$.

Let $F(U)$ denote the set of all fuzzy similarity relations from U to U. For any $\widetilde{R}_{B_1}, \widetilde{R}_{B_2} \in F(U)$, we have

1) $\widetilde{R}_{B_1}(u_i, u_j) \leq \widetilde{R}_{B_2}(u_i, u_j) \Rightarrow \widetilde{R}_{B_1} \subseteq \widetilde{R}_{B_2}$;
2) $(\widetilde{R}_{B_1} \cap \widetilde{R}_{B_2})(u_i, u_j) = \min\left\{\widetilde{R}_{B_1}(u_i, u_j), \widetilde{R}_{B_2}(u_i, u_j)\right\}$;
3) $(\widetilde{R}_{B_1} \cup \widetilde{R}_{B_2})(u_i, u_j) = \max\left\{\widetilde{R}_{B_1}(u_i, u_j), \widetilde{R}_{B_2}(u_i, u_j)\right\}$.

Based on the fuzzy relation, the definition of the fuzzy rough set FRS proposed by Dubois and Prade [35] is as follows.

Definition 2 *Let $\widetilde{R}_B$ be a fuzzy equivalence relation on U. For any fuzzy set $\chi \in F(U)$, the lower and upper approximations of χ are a pair of fuzzy sets whose membership functions are defined as*

$$\underline{R_B}\chi\,(x) = \inf_{y \in U} \max\left\{1 - \widetilde{R}_B\,(x, y)\,, \chi(y)\right\}, \tag{2}$$

$$\overline{R_B}\chi\,(x) = \sup_{y \in U} \min\left\{\widetilde{R}_B\,(x, y)\,, \chi(y)\right\}. \tag{3}$$

Definition 3 *Let $B \subseteq C$, then the fuzzy granular structure induced by the fuzzy similarity relation $\widetilde{R}_B$ is defined as*

$$G_U(\widetilde{R}_B) = \left\{[u_1]_{\widetilde{R}_B}, [u_2]_{\widetilde{R}_B}, \cdots, [u_n]_{\widetilde{R}_B}\right\}, \tag{4}$$

where $[u_i]_{\widetilde{R}_B} = \left(\widetilde{R}_B(u_i, u_1), \widetilde{R}_B(u_i, u_2), \cdots, \widetilde{R}_B(u_i, u_n)\right)$ is called the fuzzy granule of the sample u_i w.r.t. $\widetilde{R}_B$. $G_U(\widetilde{R}_B)$ can be denoted as the fuzzy matrix $M_{\widetilde{R}_B} = \left[\widetilde{R}_B\left(u_i, u_j\right)\right]_{n\times n}$.

Wang et al. proposed a fuzzy rough entropy model in the fuzzy approximation space in [17]. The proposed fuzzy rough entropy takes advantage of the rough entropy while compensating for the limitation that the rough entropy is only applicable to nominal attribute data.

Definition 4 *The fuzzy rough entropy w.r.t. the fuzzy similarity relation $\widetilde{R}_B$ is denoted as*

$$E(B) = E\left(\widetilde{R}_B\right) = -\frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{1}{\left|[u_i]_{\widetilde{R}_B}\right|}, \tag{5}$$

where $|[u_i]_{\widetilde{R}_B}| = \sum_{j=1}^{n} \widetilde{R}_B(u_i, u_j)$.

Regarding the calculation of $[u_i]_{\widetilde{R}_B}$ in the above calculation formula, the commonly used method is the direct intersection method, i.e., $[u_i]_{\widetilde{R}_B} = \cap_{b\in B}[u_i]_{\widetilde{R}_b}$. However, in high-dimensional datasets, the difference of the membership degree to fuzzy similarity relations may be very small due to multiple intersection operations. Thus the fuzzy rough entropy obtained from fuzzy similarity relations will not really reflect the relation between samples. Here is an example to illustrate the problem with intersection operations.

Example 1 *As shown in Table 1, it shows a fuzzy information system, where $U = \{u_1, u_2, u_3, u_4, u_5, u_6\}$ and $C = \{c_1, c_2, c_3\}$. In the attribute set C, c_1 is a nominal attribute, while c_2 and c_3 are numerical attributes. Firstly, the attribute values of c_2 and c_3 need to be preprocessed using the* min − max *normalization method. The preprocessed results are shown on the right side of Table 1.*

Table 1. Initial and standardized fuzzy information system

U	c_1	c_2	c_3	c_1	c_2	c_3
u_1	C	0.9	3	C	1	$\frac{1}{7}$
u_2	B	0.2	5	B	$\frac{1}{8}$	$\frac{3}{7}$
u_3	C	0.1	2	C	0	0
u_4	A	0.4	7	A	$\frac{3}{8}$	$\frac{5}{7}$
u_5	C	0.6	8	C	$\frac{5}{8}$	$\frac{6}{7}$
u_6	A	0.3	9	A	$\frac{1}{4}$	1

For any $c_i \in C$, let $\widetilde{R}_{c_i}$ denote the fuzzy similarity relation caused by c_i. In order to calculate the fuzzy rough entropy w.r.t. $\widetilde{R}_C$, we need to calculate the fuzzy similarity between samples on C. We take the fuzzy similarities $\widetilde{R}_C(u_2, u_4)$ and $\widetilde{R}_C(u_2, u_5)$ as an example. To get $\widetilde{R}_C(u_2, u_4)$,we first calculate the fuzzy similarity under each attribute as $\widetilde{R}_{c_1}(u_2, u_4) = 1 - |B - A| = 0$, $\widetilde{R}_{c_2}(u_2, u_4) = 1 - |\frac{1}{8} - \frac{3}{8}| = 0.7500$, $\widetilde{R}_{c_3}(u_2, u_4) = 1 - |\frac{3}{7} - \frac{5}{7}| \approx 0.7143$. Next, according to the intersection operation, we get the fuzzy similarity under the attribute set C as $\widetilde{R}_C(u_2, u_4) = \min\left\{\widetilde{R}_{c_1}(u_2, u_4), \widetilde{R}_{c_2}(u_2, u_4), \widetilde{R}_{c_3}(u_2, u_4)\right\} = 0$. In the same way, we calculate that $\widetilde{R}_{c_1}(u_2, u_5) = 1 - |B - C| = 0$, $\widetilde{R}_{c_2}(u_2, u_5) = 1 - |\frac{1}{8} - \frac{5}{8}| = 0.5000$, $\widetilde{R}_{c_3}(u_2, u_5) = 1 - |\frac{3}{7} - \frac{6}{7}| \approx 0.5714$, $\widetilde{R}_C(u_2, u_5) = \min\left\{\widetilde{R}_{c_1}(u_2, u_5), \widetilde{R}_{c_2}(u_2, u_5), \widetilde{R}_{c_3}(u_2, u_5)\right\} = 0$. It is shown that both fuzzy similarities have the same similarity of 0, which determined only by the nominal attribute c_1. We can conclude that the intersection operation results in very little differentiation between similarities. In reality, $\widetilde{R}_C$ should consider the similarity of each attribute collectively, rather than paying attention only to the attribute with the smallest similarity.

From the above example, we can see that the fuzzy rough entropy estimated by the intersection operation tends to lose a lot of important uncertainty information in the high-dimensional space. Therefore, it may not effectively reflect the similarity between objects. This is the motivation for considering the improvement of fuzzy rough entropy in this study.

It is common that distance metrics are often used to describe the dissimilarity between samples in most learning algorithms. Since it does not use the intersection operation to compute the fuzzy similarity, the fuzzy similarity relation

defined by the distance metrics can compensate for the deficiency caused by the intersection operation. Therefore, it can be proved reasonable that we use the fuzzy similarity relation based on the distance metrics to construct the fuzzy rough entropy and then build the detection model to detect outliers. The next section will introduce in detail the distance-based fuzzy rough entropy proposed in this paper.

4. Distance-based fuzzy rough entropy and its related measures

In response to the shortage of existing fuzzy rough entropy, a definition of distance-based fuzzy rough entropy is proposed in this section. On the basis of fuzzy rough entropy, associated measures such as fuzzy rough joint entropy, fuzzy rough conditional entropy, and fuzzy rough mutual information are discussed. In addition, we examine several properties of these fuzzy rough measures.

4.1. Distance-based fuzzy rough entropy

Considering the need to effectively measure the difference between numerical and nominal attribute values simultaneously, a hybrid distance metric is proposed as follows.

Definition 5 *For the mixed attribute set B, which contains both nominal and numerical attributes, we partition B into a nominal attribute subset $B_1 = \{c_{p_1}, c_{p_2}, \cdots, c_{p_{|B_1|}}\}$ and a numerical attribute subset $B_2 = \{c_{q_1}, c_{q_2}, \cdots, c_{q_{|B_2|}}\}$. It obviously satisfies $B = B_1 \cup B_2$ and $B_1 \cap B_2 = \emptyset$. For any $u_i, u_j \in U$, the hybrid distance metric $Dis_B(u_i, u_j)$ between u_i and u_j w.r.t. B is computed by*

$$Dis_B(u_i, u_j) = Dis_{B_1}(u_i, u_j) + Dis_{B_2}(u_i, u_j), \tag{6}$$

where Dis_{B_1} denotes the normalized Hamming distance, which measures the difference between the values of nominal attributes, and Dis_{B_2} denotes the Euclidean distance, which measures the difference between numerical attribute values. If $B_1 = \emptyset$, then $Dis_B(u_i, u_j) = Dis_{B_2}(u_i, u_j)$. Similarly, if $B_2 = \emptyset$, then $Dis_B(u_i, u_j) = Dis_{B_1}(u_i, u_j)$.

To better process nominal, numerical, or mixed attribute data, this subsection constructs a hybrid fuzzy similarity, which is defined as follows.

Definition 6 *For any $B \subseteq C, u_i, u_j \in U$, the hybrid fuzzy similarity $\widetilde{R}_B(u_i, u_j)$ between u_i and u_j w.r.t. B is defined by*

$$\widetilde{R}_B\left(u_i, u_j\right) = \begin{cases} 1 - \frac{Dis_B(u_i, u_j)}{|B|}, & Dis_B\left(u_i, u_j\right) \leq (1-\delta)|B|; \\ 0, & Dis_B\left(u_i, u_j\right) > (1-\delta)|B|, \end{cases} \tag{7}$$

where δ is the adjustable threshold value, and $\delta \in [0, 1]$.

Similarly, it can be derived that $\widetilde{R}_B(u_i, u_j)$ takes values in the range of $[0, 1]$. Furthermore, it can be analyzed that if the threshold δ is set larger, the more likely the fuzzy similarity will be reset to zero. In Section 6, We will introduce the use of the parameter δ in the experiment preparation and analyze in detail the sensitivity of the model to the parameter in the last subsection.

Example 2 *Continued Example 1. By Definition 5, to compute the hybrid distance metric $Dis_C(u_2, u_4)$ and $Dis_C(u_2, u_5)$, we partition B into $B_1 = \{c_1\}$ and $B_2 = \{c_2, c_3\}$. Then $Dis_C(u_2, u_4) = Dis_{B_1}(u_2, u_4) + Dis_{B_2}(u_2, u_4) = \left(f_{c_1}(u_2) \oplus f_{c_1}(u_4)\right) + \sqrt{\sum_{k=1}^{2}\left(f_{c_k}(u_2) - f_{c_k}(u_4)\right)^2} = 1 + \sqrt{(0.2500^2 + 0.2857^2)} \approx 1.3796$. And $Dis_C(u_2, u_5) = Dis_{B_1}(u_2, u_5) + Dis_{B_2}(u_2, u_5) = \left(f_{c_1}(u_2) \oplus f_{c_1}(u_5)\right) + \sqrt{\sum_{k=1}^{2}\left(f_{c_k}(u_2) - f_{c_k}(u_5)\right)^2} = 1 + \sqrt{(0.5000^2 + 0.4286^2)} \approx 1.6586$.*

By Definition 6, we can calculate that $\widetilde{R}_C(u_2, u_4) = 1 - \frac{Dis_C(u_2, u_4)}{|C|} = 1 - \frac{1.3796}{3} \approx 0.5401$, and $\widetilde{R}_C(u_2, u_5) = 1 - \frac{Dis_C(u_2, u_5)}{|C|} = 1 - \frac{1.6586}{3} \approx 0.4471$.

It can be analyzed that the fuzzy similarities $\widetilde{R}_C(u_2, u_4)$ and $\widetilde{R}_C(u_2, u_5)$ obtained by using the distance metric are respectively 0.5401 and 0.4471. The difference between them is not 0, which is larger than the difference in fuzzy similarity (0) obtained from the intersection operation. The results show that the distance method has a stronger distinguishing ability than the intersection operation method. Therefore, it can be justified that we use the distance-based fuzzy similarity to construct the fuzzy rough entropy which can compensate for the information loss caused by the intersection operation.

Obviously, the new fuzzy rough entropy will slow down the computation speed of fuzzy similarity relation when replacing the original intersection operation. However, it compensates for the previous fuzzy rough entropy's inability to efficiently differentiate the similarity of samples in high-dimensional space. The anomaly detection model constructed by it shows excellent performance in the subsequent experimental results. Thus, the slight slowdown in computational speed is offset by the significant improvement in performance and the ability to efficiently process data in high-dimensional spaces.

As a result, the fuzzy similarity between samples can be obtained based on the hybrid distance, which leads to the definition of fuzzy information granule as follows.

Definition 7 *Let $B \subseteq C$, then the fuzzy information granule of the sample u_i w.r.t. $\widetilde{R}_B$ is calculated as*

$$SIM_{\widetilde{R}_B}(u_i) = SIM_B(u_i) = (\widetilde{R}_B(u_i, u_1), \widetilde{R}_B(u_i, u_2), \cdots, \widetilde{R}_B(u_i, u_n)). \tag{8}$$

It can be analyzed that if $\widetilde{R}_B\left(u_i, u_j\right) = 0$, u_j may definitely not belong to the fuzzy granule $SIM_B(u_i)$. Conversely, if $\widetilde{R}_B\left(u_i, u_j\right) = 1$, then it implies that u_j must belong to $SIM_B(u_i)$. Hence, our corresponding fuzzy granular structure can be re-expressed as $G_U(\widetilde{R}_B) = G_U(B) = \{\mathrm{SIM}_B(u_1), \mathrm{SIM}_B(u_2), \ldots, \mathrm{SIM}_B(u_n)\}$.

Definition 8 *The new fuzzy rough entropy w.r.t. $\widetilde{R}_B$ is calculated as*

$$E(B) = E\left(\widetilde{R}_B\right) = -\frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{1}{|SIM_B(u_i)|}, \tag{9}$$

where $|SIM_B(u_i)| = \sum_{j=1}^{n} \widetilde{R}_B(u_i, u_j)$.

It is easy to get that for any $u_i, u_j \in U$, if $\widetilde{R}_B(u_i, u_j) = 1, |\mathrm{SIM}_B(u_i)| = n$, so we have $E(B) = \log_2 n$. On the contrary, for any $u_i \neq u_j$, if $\widetilde{R}_B(u_i, u_j) = 0, |\mathrm{SIM}_B(u_i)| = 1$, then we have $E(B) = 0$.

4.2. Related measures and their properties

This subsection specifies the relevant measures of fuzzy rough joint entropy, fuzzy rough conditional entropy, and fuzzy rough mutual information with their properties.

Definition 9 *For any $B_1, B_2 \subseteq C$, the fuzzy rough joint entropy of B_1 and B_2 is given by*

$$E(B_1, B_2) = -\frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{1}{\left|SIM_{B_1}(u_i) \cap SIM_{B_2}(u_i)\right|}. \tag{10}$$

Proposition 1 *If $B_1, B_2 \subseteq C$, then $E(B_1, B_2) \leq \min\{E(B_1), E(B_2)\}$.*

Proof: *$SIM_{B_1}(u_i) \cap SIM_{B_2}(u_i) = \{(\widetilde{R}_{B_1} \cap \widetilde{R}_{B_2})(u_i, u_1), \cdots, (\widetilde{R}_{B_1} \cap \widetilde{R}_{B_2})(u_i, u_n)\} = \{\min\{\widetilde{R}_{B_1}(u_i, u_1), \widetilde{R}_{B_2}(u_i, u_1)\}, \cdots, \min\{\widetilde{R}_{B_1}(u_i, u_n), \widetilde{R}_{B_2}(u_i, u_n)\}\}$, then $|SIM_{B_1}(u_i) \cap SIM_{B_2}(u_i)| \leq |SIM_{B_1}(u_i)|$, and $|SIM_{B_1}(u_i) \cap SIM_{B_2}(u_i)| \leq |SIM_{B_2}(u_i)|$. As a result, by Definition 9, there is $E(B_1, B_2) \leq E(B_1), E(B_1, B_2) \leq E(B_2)$. Therefore, $E(B_1, B_2) \leq \min\{E(B_1), E(B_2)\}$.*

Proposition 1 shows that when there are two attribute subsets inducing two fuzzy granules each, their fuzzy rough joint entropy is lower than the respective fuzzy rough entropy of both. In other words, the joint uncertainty of two attribute subsets is limited by the uncertainty of each subset.

Definition 10 *Under the condition of B_1, the fuzzy rough conditional entropy of B_2 is given by*

$$E(B_2|B_1) = -\frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{\left|SIM_{B_1}(u_i)\right|}{\left|SIM_{B_1}(u_i) \cap SIM_{B_2}(u_i)\right|}. \tag{11}$$

Proposition 2 *$E(B_2|B_1) = E(B_1, B_2) - E(B_1)$.*

Proof: *$E(B_2|B_1) + E(B_1) = -\frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{|SIM_{B_1}(u_i)|}{|SIM_{B_1}(u_i) \cap SIM_{B_2}(u_i)|} - \frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{1}{|SIM_{B_1}(u_i)|} = -\frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{|SIM_{B_1}(u_i)|}{|SIM_{B_1}(u_i) \cap SIM_{B_2}(u_i)|} \frac{1}{|SIM_{B_1}(u_i)|} = -\frac{1}{n}\sum_{i=1}^{n} \log_2 \frac{1}{|SIM_{B_1}(u_i) \cap SIM_{B_2}(u_i)|} = E(B_1, B_2)$.*

Proposition 2 tells us that in order to obtain the uncertainty of a certain attribute subset under the condition that another attribute subset is known, we can calculate the fuzzy rough conditional entropy from the fuzzy rough entropy of the known attribute subset and the fuzzy rough joint entropy of both subsets.

Definition 11 *The fuzzy rough mutual information between B_1 and B_2 is defined as*

$$I(B_1;B_2) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{\left|SIM_{B_1}(u_i)\cap SIM_{B_2}(u_i)\right|}{\left|SIM_{B_1}(u_i)\right|\left|SIM_{B_2}(u_i)\right|}. \tag{12}$$

In Definition 11, the fuzzy rough mutual information $I(B_1;B_2)$ reflects the correlation between B_1 and B_2. The larger the fuzzy rough mutual information between B_1 and B_2, the higher the correlation between B_1 and B_2.

Proposition 3 $I(B_1;B_2) = E(B_2) - E(B_2|B_1) = E(B_1) - E(B_1|B_2)$.

Proof: $E(B_2) - E(B_2|B_1) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{|SIM_{B_2}(u_i)|} + \frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|SIM_{B_1}(u_i)|}{|SIM_{B_1}(u_i)\cap SIM_{B_2}(u_i)|} = \frac{1}{n}\sum_{i=1}^{n}(\log_2\frac{|SIM_{B_1}(u_i)|}{|SIM_{B_1}(u_i)\cap SIM_{B_2}(u_i)|} - \log_2\frac{1}{|SIM_{B_2}(u_i)|}) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|SIM_{B_1}(u_i)\cap SIM_{B_2}(u_i)|}{|SIM_{B_1}(u_i)||SIM_{B_2}(u_i)|}$.

Proposition 4 $I(B_1;B_2) = I(B_2;B_1) = E(B_1) + E(B_2) - E(B_1,B_2)$.

Proof: $E(B_1) + E(B_2) - E(B_1,B_2) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{|SIM_{B_1}(u_i)|} - \frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{|SIM_{B_2}(u_i)|} + \frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{|SIM_{B_1}(u_i)\cap SIM_{B_2}(u_i)|} = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{|SIM_{B_1}(u_i)||SIM_{B_2}(u_i)|} + \frac{1}{n}\sum_{i=1}^{n}\log_2\frac{1}{|SIM_{B_1}(u_i)\cap SIM_{B_2}(u_i)|} = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|SIM_{B_1}(u_i)\cap SIM_{B_2}(u_i)|}{|SIM_{B_1}(u_i)||SIM_{B_2}(u_i)|} = I(B_1;B_2) = I(B_2;B_1)$.

Propositions 3 and 4 show that the fuzzy rough mutual information measures the correlation between two subsets of attributes. It can be calculated by the fuzzy rough entropy of a certain known attribute subset minus the fuzzy rough conditional entropy for another attribute subset under the condition of the known attribute subset, or the sum of the fuzzy rough entropies of both minus the fuzzy rough joint entropy of both.

5. Anomaly detection using distance-based fuzzy rough entropy

In this section, we proceed to propose an anomaly detection model, followed by the design of the associated detection algorithm. Then its time complexity is examined. Finally, an example is shown to demonstrate how the algorithm is computed.

5.1. Detection Model

Based on the distance-based fuzzy rough entropy proposed in the previous section, we propose a concept of relative entropy, which can reflect the uncertainty of each sample more effectively.

Definition 12 *For any $u_i \in U$, let $G_{U-\{u_i\}}(B) = \{SIM_B(u_{i_1}), SIM_B(u_{i_2}), \cdots, SIM_B(u_{i_{n-1}})\}$ denote the set of fuzzy information granules of all the samples in $U-\{u_i\}$ w.r.t. $\widetilde{R}_B$. Then the relative fuzzy rough entropy of the sample u_i based on the fuzzy relation $\widetilde{R}_B$ is defined as*

$$RE_B(u_i) = \begin{cases} 1 - \frac{E_{u_i}(\widetilde{R}_B)}{E(\widetilde{R}_B)}, & E_{u_i}(\widetilde{R}_B) < E(\widetilde{R}_B); \\ 0, & otherwise, \end{cases} \tag{13}$$

where $E_{u_i}(\widetilde{R}_B) = -\frac{1}{|U-\{u_i\}|}\sum_{l=1}^{n-1}\log_2\frac{1}{|SIM_B(u_{i_l})|}$, denotes the fuzzy rough entropy under $\widetilde{R}_B$ after removing the sample u_i from U.

The degree of abnormality of u_i can be examined using $RE_B(u_i)$. When u_i is removed from the sample set U, if the value of $E_{u_i}(\widetilde{R}_B)$ is significantly less compared to $E(\widetilde{R}_B)$, it indicates that u_i is less anomalous and less likely to be an outlier. Conversely, if $E_{u_i}(\widetilde{R}_B)$ decreases very little or even increases, it indicates that u_i has a high degree of an anomaly and is more likely to be an outlier. Therefore, the lower the relative fuzzy rough entropy $RE_B(u_i)$ of u_i represents the higher the degree of the anomaly of u_i.

Additionally, the weight function formulation may efficiently aid in enlarging the difference between the anomalies and normal points. The weight function used in this paper is defined as follows.

Definition 13 *The weight function of the object u w.r.t. the fuzzy similarity relation* $\widetilde{R}_B$ *is computed by*

$$W_B(u) = \sqrt[3]{\frac{|SIM_B(u)|}{|U|}}. \tag{14}$$

Based on the idea of [32], we give a sample point greater weight when it has a closer resemblance to other sample points under a particular attribute subset B. Because this can laterally reflect that it is more consistent with the value domain distribution on that subset. On the contrary, when an object has little similarity with other objects, it is a side reflection that it deviates more from the value domain distribution on that attribute subset. And it is more likely to be an outlier in that sample set, so we will give it less weight. Furthermore, we can get $W_B(u) \in (0, 1]$.

From Definitions 12 and 13, we can get different relative entropies and weights by different similarity relations. Thus we can obtain more information about the data by computing $RE_B(u)$ and $W_B(u)$ for each object using $2^{|C|}$ hybrid similarity relations. However, the time complexity of computing all these relations will grow exponentially on $|C|$, which is obviously infeasible. Therefore, we construct a sequence of attributes and a sequence of attribute subsets to compute $RE_B(u)$ and $W_B(u)$, which is able to reduce the amount of computation.

Definition 14 *The sequence of attributes* S *is constructed as*

$$S = \langle c'_1, c'_2, \cdots, c'_m \rangle, \tag{15}$$

where $E\left(c'_j\right) \leq E\left(c'_{j+1}\right), 1 \leq j < m.$

Next, the attribute c'_1 is taken as the first set, then each time add the attribute from S to it by the forward way to acquire a new set. The operation ends when all the attributes are finally joined, at which point the new set acquired is the attribute set C. As a result, we may generate a sequence of attribute subsets.

Definition 15 *The sequence of attribute subsets* AS *is constructed as*

$$AS = \langle C_1, C_2, \cdots, C_m \rangle, \tag{16}$$

where $C_j \subseteq C, C_1 = \{c'_1\}, C_m = C,$ *and* $C_{j+1} = C_j \cup \{c'_{j+1}\}, 1 \leq j < m.$

When dealing with complex data, a single granularity may not fully capture the potential anomalous information. Each attribute or subset of attributes in the attribute sequence and attribute set sequence can induce a fuzzy information granule. By analyzing these different fuzzy information granules, we can observe and analyze the changes in the sample information at different granularities and levels, which can reveal the potential anomalous information. Finally, we can fuse the anomalous information obtained at multiple granularities to achieve more effective anomaly detection.

Then, we construct two relative entropy-based matrices and their weight matrices according to the above two sequences.

Definition 16 *The relative entropy-based matrices of these two sequences are constructed as*

$$REM_S = \left[RE_{c'_j}(u_i)\right]_{n\times m}; \tag{17}$$

$$REM_{AS} = \left[RE_{C_j}(u_i)\right]_{n\times m}, \tag{18}$$

where REM_S *refers to the relative entropy-based matrix of the attribute sequence, and* $RE_{c'_j}(u_i)$ *stands for the relative fuzzy rough entropy of the sample* u_i *based on the attribute* c'_j*. Similarly,* REM_{AS} *refers to the relative entropy-based matrix of the sequence of attribute subsets, and* $RE_{C_j}(u_i)$ *stands for the relative fuzzy rough entropy of the sample* u_i *based on the attribute subset* C_j*.*

Definition 17 *The weight matrices of these two sequences are constructed as*

$$WM_S = \left[W_{c'_j}(u_i)\right]_{n\times m}; \tag{19}$$

$$WM_{AS} = \left[W_{C_j}(u_i)\right]_{n\times m}, \tag{20}$$

where W_S *refers to the weight matrix of the attribute sequence, and* $W_{c'_j}(u_i)$ *stands for the weight of the sample* u_i *based on the attribute* c'_j*. Similarly,* W_{AS} *refers to the weight matrix of the sequence of attribute subsets, and* $W_{C_j}(u_i)$ *stands for the weight of the sample* u_i *based on the attribute subset* C_j*.*

Definition 18 *The anomaly score of object v_i based on fuzzy rough entropy is computed as*

$$score\,(u_i) = 1 - \frac{\sum_{j=1}^{|C|} \left[\left(REM_S\,(i,j) + REM_{AS}\,(i,j)\right)\left(W_S\,(i,j) + W_{AS}\,(i,j)\right)\right]}{4\,|C|}, \tag{21}$$

where $REM_S(i,j)$ and $REM_{AS}(i,j)$ refer to the elements in the ith row and jth column of the relative entropy-based matrix for the attribute sequence and attribute set sequence, respectively. Meanwhile, $W_S(i,j)$ and $W_{AS}(i,j)$ represent the elements in the ith row and jth column of the weight matrix for the attribute sequence and attribute set sequence, respectively.

From Definition 18, we can get that the anomaly score is obtained by multiplying the mean of the relative entropy of each of the two sequences with the mean of corresponding weights and then subtracting them from 1. According to Definitions 12 and 13, both the fuzzy rough relative entropy and the weight are inversely proportional to the degree of the anomaly of the sample. Therefore, it can be analyzed that the larger the anomaly score, the greater the degree of the anomaly of the sample. Finally, we set a threshold ε to determine whether a sample is abnormal or not. For any sample $u_i \in U$, if $score(u_i) > \varepsilon$, sample u_i is deemed to be an outlier in U.

5.2. Detection algorithm

This subsection gives the pseudo-code of the detection model constructed in the previous subsection and analyzes the time complexity of the algorithm.

Algorithm 1: FREAD algorithm

Input: Fuzzy Information System $FIS = (U, C, V, f)$, parameter δ, threshold ε
Output: Outlier Set (OS)

1 $OS = \emptyset$;
2 **for** $j = 1$ *to* m **do**
3 Compute the fuzzy similarity relation matrix $M_{\widetilde{R}_{c_j}}$ by Definitions 5 and 6 ;
4 Compute the fuzzy rough entropy $E(c_j)$;
5 **for** $i = 1$ *to* n **do**
6 Compute the weight $W_{c_j}(u_i)$;
7 Compute the fuzzy rough relative entropy $RE_{c_j}(u_i)$;
8 **end**
9 **end**
10 Construct the sequence of attributes $S = \{c'_1, c'_2, \cdots, c'_m\}$;
11 Construct the sequence of attribute subsets $AS = \{C_1, C_2, \cdots, C_m\}$;
12 **for** $j = 1$ *to* m **do**
13 Compute the fuzzy similarity relation matrix $M_{\widetilde{R}_{C_j}}$ by Definitions 5 and 6 ;
14 Compute the fuzzy rough entropy $E(C_j)$;
15 **for** $i = 1$ *to* n **do**
16 Compute the weight $W_{C_j}(u_i)$;
17 Compute the fuzzy rough relative entropy $RE_{C_j}(u_i)$;
18 **end**
19 **end**
20 **for** $i = 1$ *to* n **do**
21 Calculate $score(u_i)$ by Definition 18;
22 **if** $score(u_i) > \varepsilon$ **then**
23 $OS = OS \cup \{u_i\}$;
24 **end**
25 **end**
26 **return** OS.

In Algorithm 1, OS is initially set to $\emptyset$. In the "for" loop of Steps 2-9, Step 3 calculates the fuzzy similarity matrix for each attribute, and the number of cycles is n^2 for each outer loop. Steps 4-8 are to calculate the relative fuzzy rough

entropy and weight for each sample under each attribute, and the number of cycles is also n^2 for each outer loop. Thus, the number of cycles in Steps 2-9 is mn^2. Step 10 is a fast sorting of attributes according to the size of their fuzzy rough entropy, and its complexity is generally $O(n \log n)$. In the "for" loop of Steps 12-19, Step 13 is to compute the fuzzy similarity matrix for each attribute subset, and the number of loops is $|C_j| \times n \times n$ for each outer loop, where $|C_j|$ denotes the size of the jth attribute subset. And in the mth outer loop, the size of the attribute subset is m, then the number of loops is $m \times n^2$. Steps 14-18 are to calculate the fuzzy rough relative entropy and weight of each sample under each attribute subset from AS, and the number of loops is n^2 for each outer loop. Thus Steps 12-19 have a loop count of $\sum_{k=1}^{m} k \times n^2 = \frac{m(m+1)}{2} n^2$. Steps 19-23 calculate the outlier score for each sample and determine if it is an outlier with a loop count of mn. According to the above analysis, the total number of cycles is $mn^2 + n \log n + \frac{m(m+1)}{2} n^2 + mn$. Therefore, in the worst case, the time complexity of Algorithm 1 is $O(m^2 n^2)$.

5.3. Detection example

In this subsection, an example of a fuzzy information system is used to specify the study content of the first two subsections above.

Example 3 *Continued Example 2. The distance matrix of each attribute is calculated by Definition 5. And by Definition 6, we set the parameter δ to 0.5, then the similarity matrix of each attribute $M_{\widetilde{R}_{c_i}}$ is calculated separately from the corresponding distance matrix, which is as follows.*

$$M_{\widetilde{R}_{c_1}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad M_{\widetilde{R}_{c_2}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0.6250 & 0 \\ 0 & 1 & 0.8750 & 0.7500 & 0.5000 & 0.8750 \\ 0 & 0.8750 & 1 & 0.6250 & 0 & 0.7500 \\ 0 & 0.7500 & 0.6250 & 1 & 0.7500 & 0.8750 \\ 0.6250 & 0.5000 & 0 & 0.7500 & 1 & 0.6250 \\ 0 & 0.8750 & 0.7500 & 0.8750 & 0.6250 & 1 \end{bmatrix},$$

$$M_{\widetilde{R}_{c_3}} = \begin{bmatrix} 1 & 0.7143 & 0.8571 & 0 & 0 & 0 \\ 0.7143 & 1 & 0.5714 & 0.7143 & 0.5714 & 0 \\ 0.8571 & 0.5714 & 1 & 0 & 0 & 0 \\ 0 & 0.7143 & 0 & 1 & 0.8571 & 0.7143 \\ 0 & 0.5714 & 0 & 0.8571 & 1 & 0.8571 \\ 0 & 0 & 0 & 0.7143 & 0.8571 & 1 \end{bmatrix}.$$

By Definition 8, the fuzzy rough entropy of each attribute is calculated separately using the similarity matrix as $E(c_1) \approx 1.1258, E(c_2) \approx 1.5457, E(c_3) \approx 1.7088$.

By Definition 14, the sequence of attributes is constructed by sorting the attributes according to the size of the obtained fuzzy rough entropy. And then the sequence of attribute subsets is also constructed by Definition 15. The two sequences are as follows.

$S = \langle c_1, c_3, c_2 \rangle, \quad AS = \langle C_1, C_2, C_3 \rangle = \langle \{c_1\}, \{c_1, c_3\}, \{c_1, c_2, c_3\} \rangle.$

By Definition 6, the parameter value δ remains 0.5, and the similarity matrix of each attribute subset $M_{\widetilde{R}_{C_i}}$ in the sequence of attribute subsets AS is calculated as follows.

$$M_{\widetilde{R}_{C_1}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad M_{\widetilde{R}_{C_2}} = \begin{bmatrix} 1 & 0 & 0.9286 & 0 & 0.6429 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0.9286 & 0 & 1 & 0 & 0.5714 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0.8571 \\ 0.6429 & 0 & 0.5714 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.8571 & 0 & 1 \end{bmatrix},$$

$$M_{\widetilde{R}_{C_3}} = \begin{bmatrix} 1 & 0 & 0.6633 & 0 & 0.7311 & 0 \\ 0 & 1 & 0.5179 & 0.5401 & 0 & 0 \\ 0.6633 & 0.5179 & 1 & 0 & 0.6464 & 0 \\ 0 & 0.5401 & 0 & 1 & 0.5707 & 0.8960 \\ 0.7311 & 0 & 0.6464 & 0.5707 & 1 & 0.5329 \\ 0 & 0 & 0 & 0.8960 & 0.5329 & 1 \end{bmatrix}.$$

By Definition 16, the relative entropy-based matrix of each sequence is calculated in turn, which is as follows.

$$REM_S = \begin{bmatrix} 0.2894 & 0.0992 & 0 \\ 0 & 0.2299 & 0.1884 \\ 0.2894 & 0.0739 & 0.1033 \\ 0.1553 & 0.1809 & 0.1862 \\ 0.2894 & 0.1867 & 0.1789 \\ 0.1553 & 0.0785 & 0.2002 \end{bmatrix}, \quad REM_{AS} = \begin{bmatrix} 0.2894 & 0.3399 & 0.0813 \\ 0 & 0 & 0.0294 \\ 0.2894 & 0.3125 & 0.1800 \\ 0.1553 & 0.1816 & 0.2180 \\ 0.2894 & 0.2136 & 0.2762 \\ 0.1553 & 0.1816 & 0.0877 \end{bmatrix}.$$

By Definition 17, the weight matrix of each sequence is calculated respectively as follows.

$$WM_S = \begin{bmatrix} 0.7937 & 0.6470 & 0.7539 \\ 0.5503 & 0.8736 & 0.8412 \\ 0.7937 & 0.8152 & 0.7397 \\ 0.6934 & 0.8736 & 0.8181 \\ 0.7937 & 0.8355 & 0.8181 \\ 0.6934 & 0.8826 & 0.7539 \end{bmatrix}, \quad WM_{AS} = \begin{bmatrix} 0.7937 & 0.7539 & 0.7362 \\ 0.5503 & 0.5503 & 0.7000 \\ 0.7937 & 0.7469 & 0.7782 \\ 0.6934 & 0.6764 & 0.7943 \\ 0.7937 & 0.7173 & 0.8340 \\ 0.6934 & 0.6764 & 0.7398 \end{bmatrix}.$$

By Definition 18, the anomaly score of u_1 *is computed as* $score(u_1) \approx 0.8621$. *The same can be obtained,* $score(u_2) \approx 0.9448$, $score(u_3) \approx 0.8373$, $score(u_4) \approx 0.8630$, $score(u_5) \approx 0.8090$, $score(u_6) \approx 0.8945$.

Then, we set ε *to* 0.8700, *and compare the value of the anomaly score of each object with the threshold* ε *in turn. If the score of any object is greater than* ε, *we add it to* OS. *In the end, the samples* u_2 *and* u_6 *will be judged as anomalies and added to* OS.

Table 2. Experimental datasets

No.	Datasets	Abbr.	Conditional attributes		Outliers	Samples
			Numerical	Nominal		
1	Annealing	Annealing	9	29	42	798
2	Arrhythmia_variant1	Arrhythmia	206	73	66	452
3	Bands_band_6_variant1	Bands	20	20	6	318
4	CreditA_plus_42_variant1	Credit	6	9	42	425
5	German_1_14_variant1	German	7	13	14	714
6	Heart270_2_16_variant1	Heart	6	7	16	166
7	Hepatitis_2_9_variant1	Hepatits	6	13	9	94
8	Horse_1_12_variant1	Horse	8	19	12	256
9	Sick_sick_35_variant1	Sick	7	22	35	3756
10	Cardiotocography_2and3_3_variant1	Cardiotocography	21	0	33	1688
11	Diabetes_tested_positive_26_variant1	Diabetes	8	0	26	526
12	Ionosphere_b_24_variant1	Ionosphere	34	0	24	249
13	Iris_Irisvirginica_11_variant1	Iris	4	0	11	111
14	Pageblocks_258_variant1	Pageblocks	10	0	258	5171
15	Thyoid	Thyoid	6	0	93	3772
16	Wbc_malignant_39_variant1	Wbc	9	0	39	483
17	Wdbc_M_39_variant1	Wdbc	31	0	39	396
18	Yeast_ERL_5_variant1	Yeast	8	0	5	1141
19	Lymphography	Lymphography	0	8	6	148
20	Monks_0_25_variant1	Monks	0	6	25	253
21	Mushroom_p_85_variant1	Mushroom	0	22	85	4293
22	Vote_republican_29_variant1	Vote	0	16	29	296

6. Experiments

In this section, a total of 22 datasets including nominal attribute datasets, numerical attribute datasets, and mixed attribute datasets are gathered to evaluate the proposed algorithm and the other nine anomaly detection algorithms. The performance of the proposed algorithm is compared and analyzed with other algorithms.

6.1. Experiment preparation

First, we will discuss how the datasets, comparison algorithms, and evaluation indexes utilized in this research are prepared before the tests.

6.1.1. Datasets

The datasets used in this paper are taken from an open-source webpage[1]. Table 2 displays the particular information on the experimental datasets. There are four nominal attribute datasets, nine numerical attribute datasets, and nine mixed attribute datasets. In addition, the number of objects ranges from 94 to 5171 and the number of attributes ranges from 4 to 279. In our experiment, the min – max normalization approach is applied to transform the range of original numerical values into $[0, 1]$.

6.1.2. Compare algorithms

The outlier detection algorithms used for comparison in this article include Connectivity-based Outlier Factor (COF) [36], DIStance-based outlier detection algorithm (DIS) [37], INFluenced Outlierness (INFLO) [38], Local Distance-based Outlier Factor (LDOF) [39], Local Outlier Probability (LoOP)[40], a joint learning-based MIXed-type anomaly detection method (MIX) [41], Outlier Detection using Indegree Number (ODIN) [42], Self-Representation based Outlier detection (SRO) [43] and VARiance structural score (VarE) [44]. Among the above comparison algorithms, DIS and LDOF are both distance-based outlier detection algorithms. COF, INFLO and LoOP are density-based algorithms. MIX builds a joint learning framework for detecting clustered and dispersed outliers in mixed-type data. And ODIN is a k-nearest neighbor graph-based outlier detection algorithm that uses indegree numbers. SRO combines a sparse representation with random wandering on the sample graph, which is able to detect outliers in computer vision tasks. VarE is a method for detecting outliers using structural scores, which are calculated by measuring the angular variance weighted by the data representation.

In the experiments, the above algorithm models need to set certain parameters before running. Among them, for the parameter k required in LDOF, ODIN and density-based algorithms, we let k range from 1 to 60 with a step size of 1. The parameter in SRO varies from 2 to 20. We run the MIX experiment 10 times, and then take the maximum AUC value as the final result. The parameters used in the remaining comparison algorithms are set mainly based on the parameters provided in the experimental section of the relevant literature. With regard to FREAD, we let δ range from 0 to 1 by increasing 0.05 each time, and we take the experimental result with the best performance as the final result.

6.1.3. Evaluation indexes

We mainly employ ROC (Receiver Operating Characteristic) curve and AUC (Area Under Curve) to evaluate the effectiveness of the method according to Refs. [45] and [34]. The main idea is as follows: the final result of the general outlier detection method is the value of the outlier factor for each sample in U. And the larger the value, the more likely the object is an outlier. Therefore, for each outlier detection algorithm, all samples in U will be sorted in descending order according to their outlier values. Furthermore, the experimenters will typically provide a positive integer p which is no greater than the number of samples in U. The top p objects in the order will be judged as outliers.

For any p, let $OS(p)$ signify the outlier set determined by the given p, and OS_{true} indicates the set of all real outliers from the dataset. The horizontal and vertical coordinates of the ROC curve are $FPR(p)$ (False Positive Rate) and $TPR(p)$ (True Positive Rate), respectively.

$TPR(p)$ represents the rate of all correctly predicted anomalies to all true anomalies, and is calculated as

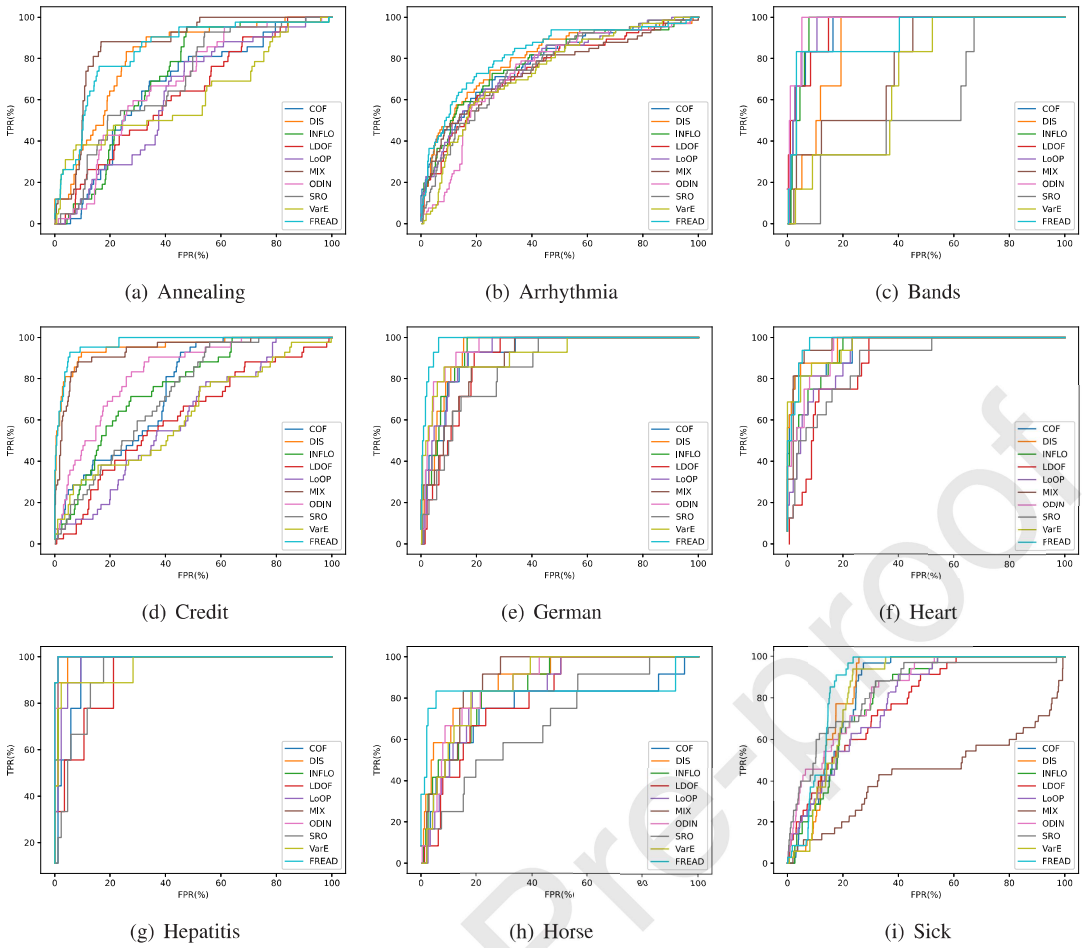
$$TPR(p) = \frac{|OS(p) \cap OS_{true}|}{|OS_{true}|} \times 100\%. \tag{22}$$

The percentage of samples projected to be anomalies but are actually normal points to all true normal points is known as $FPR(p)$, and is calculated as

$$FPR(p) = \frac{|OS(p) - OS_{true}|}{|U - OS_{true}|} \times 100\%. \tag{23}$$

When comparing the performance of outlier detection models by ROC curves, the closer the ROC curve is to the upper left corner of the first quadrant, the better the related algorithm performs. However, it can be tricky to

[1] https://github.com/BElloney/Outlier-detection



(a) Annealing (b) Arrhythmia (c) Bands

(d) Credit (e) German (f) Heart

(g) Hepatitis (h) Horse (i) Sick

Fig. 1. ROC on mixed datasets

intuitively tell which algorithm performs better when the ROC curves of two or more algorithms mostly overlap or have more cross sections. For this problem, we adopt a new performance indicator AUC that may be used to assess the performance of each algorithm. And the calculating formula for it is as follows.

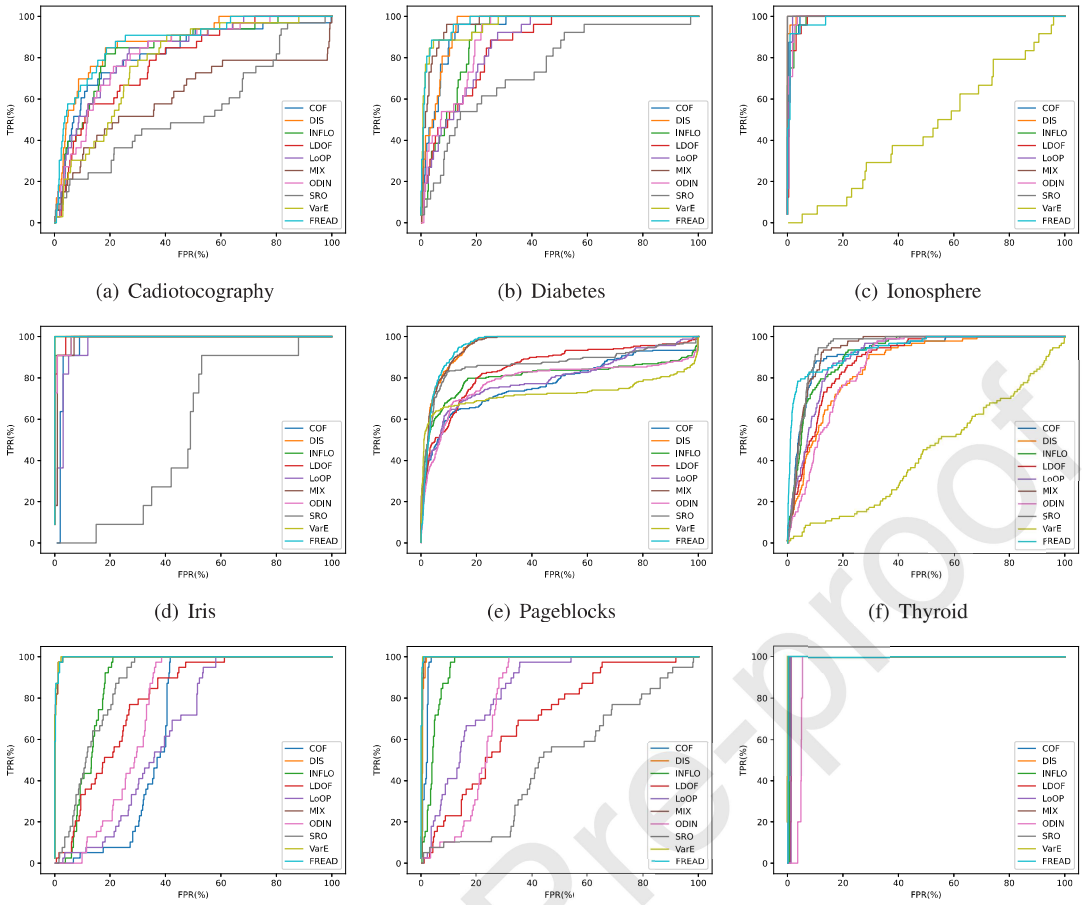
$$\mathrm{AUC} = \underset{o_i \in OS_{true}, o_j \in U - OS_{true}}{\mathrm{Mean}} \begin{cases} 1, & score(o_i) < score(o_j) \\ 0.5, & score(o_i) = score(o_j) \\ 0, & score(o_i) < score(o_j) \end{cases}. \tag{24}$$

It takes values in the range of $[0, 1]$. A greater AUC indicates improved performance.

6.2. *Experimental results on ROC curves*

This subsection examines the experimental outcomes of each method on various datasets regarding the ROC curves. Figs. 1-3 show separately the ROC curves for the mixed, numerical, and nominal attribute datasets. And the blue curve in each figure is the ROC curve of the proposed algorithm FREAD. From Figs. 1-3, we can find that the blue curves in the datasets such as Arrhythmia, Credit, German, Heart, Hepatitis, Iris, Sick, Pageblocks, Thyroid, Wdbc, and Monks are obviously closest to the upper left corner of the figures compared with the curves of other algorithms. In the datasets Annealing, Diabetes, Wbc, Yeast and Lymphography, the FREAD curve overlaps almost completely with the MIX curves, indicating that there is little difference in how well the three algorithms perform on Vote. And in the datasets Bands, Horse, Mushroom and Vote, although the ROC curve for FREAD is not the curve closest to the upper left corner, it is much nearer to the top left corner than most curves. As a result, it is clear that FREAD mostly outperforms other methods in terms of ROC.

However, the ROC curves of several algorithms contain one or more cross-sections in some datasets, such as Annealing, Cardiotocography, Diabetes, Vote, etc. The issue makes it hard to determine for sure whether the approach



(a) Cadiotocography (b) Diabetes (c) Ionosphere

(d) Iris (e) Pageblocks (f) Thyroid

(g) Wbc (h) Wdbc (i) Yeast

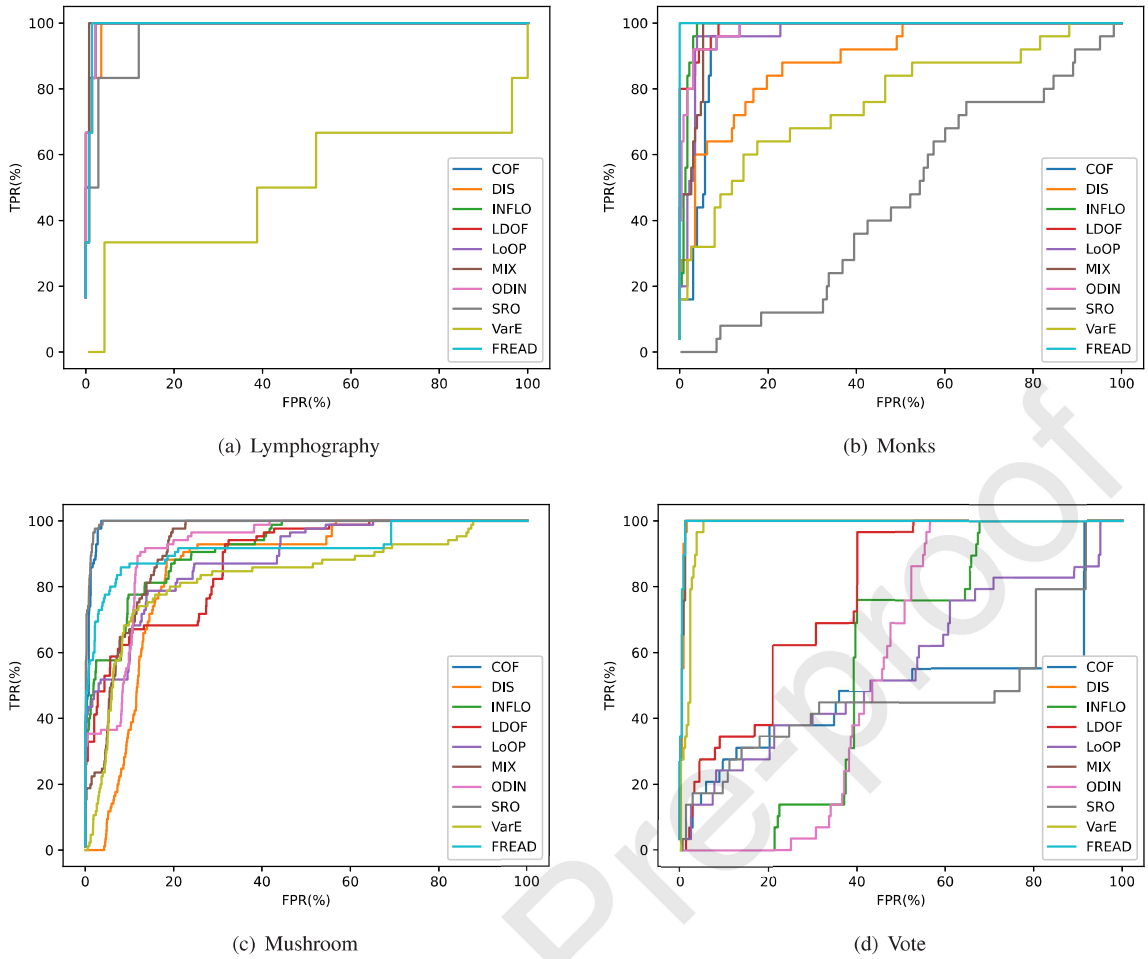
Fig. 2. ROC on numerical datasets

works better. Therefore, the next subsection will provide more experimental findings of these comparison algorithms on AUC.

6.3. Experimental results on AUC

Table 3 displays the experimental comparative outcomes for AUCs, where the bold numbers in each row mark the best performing algorithms among all the algorithms. From the analysis of Table 3, it can be concluded that FREAD shows better detection performance than the other 9 algorithms on most of the datasets. It occupies 14 best AUCs in 22 datasets, while the remaining 9 algorithms have 0, 3, 0, 0, 0, 6, 1, 3, and 3 best AUCs respectively, which are much fewer than FREAD. In addition, the average values of AUC may be used to compare algorithm performance more successfully. The average AUC values of COF, DIS, INFLO, LDOF, LoOP, MIX, ODIN, SRO, VarE, and FREAD correspond to 0.863, 0.930 0.885, 0.851, 0.845, 0.905, 0.869, 0.766, 0.803, and 0.949, respectively. The best result among them is produced by FREAD, which is much higher than that of the other methods.

For datasets having nominal attributes, algorithms like DIS, COF, and LDOF must substitute distinct integer values for the nominal attributes, which will impact how well the algorithmic model performs. In comparison, mixed data can be dealt with nicely by Definition 5 devoid of the need to convert attribute values. Thus the fuzzy similarity relation created by the hybrid distance may maintain important information to a significant extent. Moreover, the method makes full use of the effectiveness of fuzzy rough calculation theory in handling fuzzy information. As a result, this subsection may successfully demonstrate how the outlier detection performance of FREAD outperforms several mainstream outlier detection algorithms and prove the reliability of the algorithm.



(a) Lymphography (b) Monks

(c) Mushroom (d) Vote

Fig. 3. ROC on nominal datasets

6.4. *Statistical tests*

In this subsection, we evaluate the statistical significance of the algorithm results according to Refs. [45] and [34]. The AUC values obtained on all data sets corresponding to each algorithm need to be first sorted from lowest to highest and assigned ordinal values $1, 2, \ldots$, denoted by the variable r_i. If two or more algorithms obtain the same AUC values, the ordinal values are averaged to obtain the average ordinal value. Then, the Friedman test is used to analyze and determine whether each algorithm performs equally well. Suppose we compare K algorithms on N datasets, and the Friedman test is calculated as follows.

$$\begin{aligned} \tau_F &= \frac{(N-1)\tau_\chi^2}{N(K-1)-\tau_\chi^2} \text{ and} \\ \tau_\chi^2 &= \frac{12N}{K(K+1)}\left(\sum_{i=1}^{K} r_i^2 - \frac{K(K+1)^2}{4}\right). \end{aligned} \tag{25}$$

The variables τ_F are F-distributed with $(K-1)$ and $(K-1)(N-1)$ degrees of freedom, respectively. If the original hypothesis of "all algorithms have the same performance" can be rejected, it indicates a considerable variation in how well certain algorithms perform. Nemenyi's post-hoc test is applied for comparing the performance of K algorithms against each other. The test calculates the critical difference (CD) of the mean series value, which is as follows.

$$CD_\alpha = q_\alpha \sqrt{\frac{K(K+1)}{6N}}, \tag{26}$$

where q_α is the critical value of Tukey's distribution. The statement that "the two algorithms perform the same" needs to be rejected under the corresponding degree of confidence if the difference between the average ordinal values of the two algorithms is greater than the critical value domain CD.

Table 3. Experimental comparison results on AUC

Datasets	COF	DIS	INFLO	LDOF	LoOP	MIX	ODIN	SRO	VarE	FREAD
Annealing	0.662	0.808	0.705	0.627	0.623	**0.864**	0.682	0.693	0.619	0.837
Arrhythmia	0.783	0.803	0.782	0.749	0.771	0.799	0.737	0.757	0.739	**0.825**
Bands	0.950	0.890	0.959	0.955	0.964	0.772	**0.978**	0.577	0.702	0.914
Credit	0.746	0.954	0.760	0.616	0.598	0.937	0.832	0.724	0.625	**0.974**
German	0.918	0.938	0.932	0.889	0.925	0.891	0.952	0.852	0.919	**0.985**
Heart	0.957	0.970	0.945	0.880	0.923	0.974	0.949	0.880	0.962	**0.981**
Hepatitis	0.959	0.990	0.995	0.922	0.976	0.999	0.997	0.932	0.962	**0.999**
Horse	0.756	0.888	0.853	0.803	0.838	**0.893**	0.870	0.685	0.869	0.834
Sick	0.842	0.853	0.805	0.779	0.788	0.433	0.837	0.840	0.842	**0.882**
Cardiotocography	0.824	0.877	0.837	0.790	0.833	0.616	0.824	0.549	0.780	**0.883**
Diabetes	0.942	0.952	0.902	0.863	0.870	0.969	0.898	0.754	0.965	**0.976**
Ionosphere	0.994	0.998	0.991	0.988	0.993	**1.000**	0.993	**1.000**	0.456	0.989
Iris	0.971	**1.000**	0.995	0.995	0.969	0.986	0.993	0.534	**1.000**	**1.000**
Pageblocks	0.776	0.956	0.807	0.849	0.791	0.956	0.784	0.870	0.727	**0.958**
Thyroid	0.926	0.862	0.920	0.887	0.909	0.941	0.863	**0.946**	0.432	0.943
Wbc	0.654	**0.997**	0.875	0.792	0.645	**0.997**	0.734	0.870	**0.997**	**0.997**
Wdbc	0.984	0.995	0.954	0.702	0.839	0.996	0.786	0.489	0.997	**0.999**
Yeast	0.997	**1.000**	0.992	0.987	0.988	**1.000**	0.954	0.995	**1.000**	0.998
Lymphography	0.994	0.989	0.994	0.992	0.993	**0.998**	0.994	0.969	0.507	0.993
Monks	0.953	0.892	0.987	0.989	0.970	0.978	0.975	0.462	0.763	**1.000**
Mushroom	0.992	0.853	0.918	0.878	0.885	0.922	0.920	**0.996**	0.830	0.921
Vote										**0.996**
Average										**0.949**

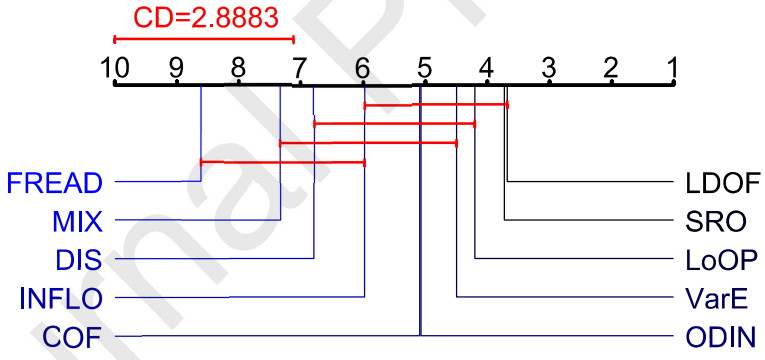


Fig. 4. Nemenyi's test plot on AUC

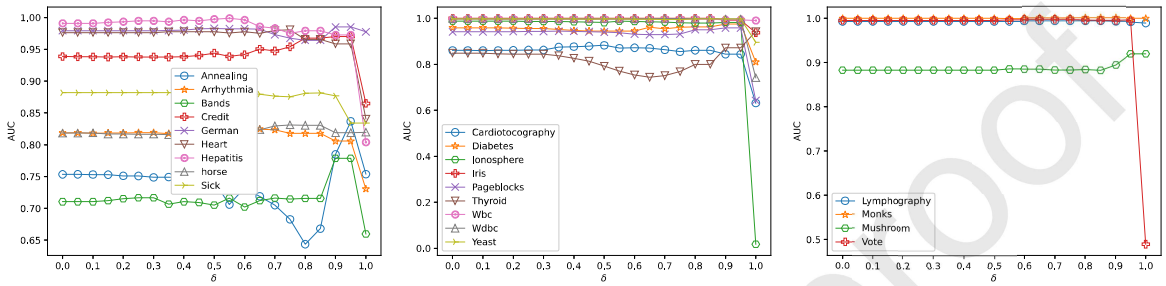
Nemenyi's test plot is constructed using the ordinal values and CD values of the algorithms. In this plot, the average ordinal value of each algorithm is represented by a dot, and the magnitude of CD is shown by a horizontal line. If a group of algorithms is connected by a horizontal line, we can conclude that there is no statistically significant difference among these algorithms.

Corresponding to the experimental part of this paper, we can get $N = 22$ as well as $K = 10$, then τ_F obeys the F distribution with degrees of freedom of 9 and 189, respectively. According to the Friedman test, when $\alpha = 0.05$, the value of τ_F is 8.8553 greater than the critical value of 1.9297, thus the original hypothesis that "all algorithms perform the same" is denied. This means that the performance of all outlier detection algorithms varies significantly.

Let significance level $\alpha = 0.05$, then $CD_{0.05} = 2.8883$. Finally, Nemenyi's test plot on AUC is shown in Fig. 4. it, we can see that the algorithm FREAD is not horizontally connected to COF, ODIN, VarE, LoOP, SRO and LDOF, which indicates that FREAD is statistically significantly different from these algorithms. However, it cannot be proved that FREAD has statistical difference with MIX, DIS and INFLO.

6.5. *Parameter sensitivity analysis*

In the last experimental session, we analyze the sensitivity of FREAD to the parameter δ. The variation curves of AUC w.r.t. the parameter δ are plotted separately for the mixed, numerical, and nominal attribute datasets in Figs. 5(a)-5(c). It can be seen from Fig. 5(a) that most of the datasets exhibit a small fluctuation of the AUC curve to the change of δ initially. While the AUC has a tendency to rise sharply to fall rapidly when δ is close to 1, such as dataset Annealing, Bands, and Sick. In Figs. 5(b)-5(c), the AUC values of all data sets generally change slightly when δ is taken in the interval $[0, 0.8]$, while the AUC curves also produce sharp changes when δ is in the interval $[0.9, 1]$, such as dataset Vowels, Ionosphere, Mushroom, and Vote. Based on the above analysis, it can be concluded that the performance of FREAD is generally less sensitive to δ in the interval $[0, 0.8]$, but more sensitive to δ in the interval $[0.9, 1]$.



(a) Mixed (b) Numerical (c) Nominal

Fig. 5. Variation curve of AUC with parameter δ

7. Conclusion

This paper proposes a distance-based fuzzy rough entropy and discusses its associated measures. The proposed new entropy makes up for the shortcomings of the existing fuzzy rough entropy using intersection operation, and can better portray the similarity between two samples in the high-dimensional space. The proposed fuzzy rough entropy is further utilized to construct an anomaly detection model, which employs two kinds of attribute sequences to mine each level of information from the multi-granular perspective. It takes full advantage of the fuzzy rough theory to deal with fuzzy information effectively and can handle multiple attribute-type data sets. Both the comparison test and statistical analysis results show that the model is more comprehensive and effective in anomaly detection. However, anomalies usually have multiple types, i.e., they are categorized as global, contextual, and collective anomalies. Different types of anomalies correspond to different application scenarios. The anomaly detection model in this paper only targets global outliers, so the performance will not meet the demand when applied to detecting contextual or collective anomalies. In order to expand the applicability scenarios of our model, we will delve into contextual outlier detection methods in our upcoming research, which more comprehensively identify anomalous behaviors in data under different contexts or conditions. In addition, we will also explore collective outlier detection methods, which primarily identify clusters of outliers that usually occur as a collective. We expect that future work will better meet the anomaly detection needs in real-world applications.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62306196, 62376230, and 62076171), Sichuan Science and Technology Program (2023YFQ0020), and the Fundamental Research Funds for the Central Universities (YJ202245).

References

[1] J. Kong, W. Jiang, Q. Tian, M. Jiang, T. S. Liu, Anomaly detection based on joint spatio-temporal learning for building electricity consumption, Applied Energy 334 (2023) 120635.

[2] H. Fanai, H. Abbasimehr, A novel combined approach based on deep autoencoder and deep classifiers for credit card fraud detection, Expert Systems with Applications (2023) 119562.

[3] F. Jiang, X. Yu, D. W. Gong, J. W. Du, A random approximate reduct-based ensemble learning approach and its application in software defect prediction, Information Sciences 609 (2022) 1147–1168.
[4] H. Asgharzadeh, A. Ghaffari, M. Masdari, F. S. Gharehchopogh, Anomaly-based intrusion detection system in the internet of things using a convolutional neural network and multi-objective enhanced capuchin search algorithm, Journal of Parallel and Distributed Computing (2023).
[5] S. Mascaro, A. E. Nicholso, K. B. Korb, Anomaly detection in vessel tracks using bayesian networks, International Journal of Approximate Reasoning 55 (1) (2014) 84–98.
[6] P. J. Rousseeuw, A. M. Leroy, Robust regression and outlier detection, John wiley & sons, 2005.
[7] Z. Y. He, X. F. Xu, S. C. Deng, Discovering cluster-based local outliers, Pattern Recognition Letters 24 (9–10) (2003) 1641–1650.
[8] B. Ali, N. Azam, A. Shah, J. T. Yao, A spatial filtering inspired three-way clustering approach with application to outlier detection, International Journal of Approximate Reasoning 130 (2021) 1–21.
[9] E. M. Knorr, R. T. Ng, A unified notion of outliers: Properties and computation, in: KDD, Vol. 97, 1997, pp. 219–222.
[10] M. M. Breunig, H. P. Kriegel, R. T. Ng, J. Sander, Lof: identifying density-based local outliers, Acm Sigmod Record 29 (2) (2000) 93–104.
[11] F. Jiang, Y. F. Sui, C. G. Cao, Some issues about outlier detection in rough set theory, Expert Systems with Applications 36 (3) (2009) 4680–4687.
[12] F. Jiang, Y. F. Sui, C. G. Cao, Outlier detection using rough set theory, in: International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing, Springer, 2005, pp. 79–87.
[13] F. Jiang, G. Z. Liu, J. W. Du, Y. F. Sui, Initialization of k-modes clustering using outlier detection techniques, Information Sciences 332 (2016) 167–183.
[14] Y. M. Chen, D. Q. Miao, H. Y. Zhang, Neighborhood outlier detection, Expert Systems with Applications 37 (12) (2010) 8745–8749.
[15] Z. Yuan, X. Y. Zhang, S. Feng, Hybrid data-driven outlier detection based on neighborhood information entropy and its developmental measures, Expert Systems with Applications 112 (2018) 243–257.
[16] X. F. Tan, C. Gao, J. Zhou, J. J. Wen, Three-way decision-based co-detection for outliers, International Journal of Approximate Reasoning (2023) 108971.
[17] Z. H. Wang, H. M. Chen, Z. Yuan, X. L. Yang, P. F. Zhang, T. R. Li, Exploiting fuzzy rough mutual information for feature selection, Applied Soft Computing 131 (2022) 109769.
[18] Z. Yuan, H. M. Chen, P. F. Zhang, J. H. Wan, T. R. Li, A novel unsupervised approach to heterogeneous feature selection based on fuzzy mutual information, IEEE Transactions on Fuzzy Systems 30 (9) (2021) 3395–3409.
[19] C. Z. Wang, Y. H. Qian, W. P. Ding, X. D. Fan, Feature selection with fuzzy-rough minimum classification error criterion, IEEE Transactions on Fuzzy Systems 30 (8) (2022) 2930–2942.
[20] B. B. Sang, W. H. Xu, H. M. Chen, T. R. Li, Active anti-noise fuzzy dominance rough feature selection using adaptive k-nearest neighbors, IEEE Transactions on Fuzzy Systems (2023) 1–15doi:10.1109/TFUZZ.2023.3272316.
[21] C. Liu, Z. Yuan, B. Y. Chen, H. M. Chen, D. Z. Peng, Fuzzy granular anomaly detection using markov random walk, Information Sciences 646 (2023) 119400.
[22] J. Dai, X. Zou, Y. Qian, X. Wang, Multifuzzy β-covering approximation spaces and their information measures, IEEE Transactions on Fuzzy Systems 31 (3) (2022) 955–969.
[23] C. Z. Wang, Y. Huang, M. W. Shao, X. D. Fan, Fuzzy rough set-based attribute reduction using distance measures, Knowledge-Based Systems 164 (2019) 205–212.
[24] R. J. Bolton, D. J. Hand, Statistical fraud detection: A review, Statistical science (2002) 235–249.
[25] E. M. Knorr, R. T. Ng, V. Tucakov, Distance-based outliers: algorithms and applications, The VLDB Journal 8 (3) (2000) 237–253.
[26] F. Jiang, Y. M. Chen, Outlier detection based on granular computing and rough set theory, Applied intelligence 42 (2015) 303–322.
[27] T. Sangeetha, A. Geetha Mary, Rough set-based entropy measure with weighted density outlier detection method, Open Computer Science 12 (1) (2022) 123–133.
[28] N. Suri, M. N. Murty, G. Athithan, Detecting outliers in categorical data through rough clustering, Natural Computing 15 (3) (2016) 385–394.
[29] F. Jiang, H. B. Zhao, J. W. Du, Y. Xue, Y. J. Peng, Outlier detection based on approximation accuracy entropy, International Journal of Machine Learning and Cybernetics 10 (9) (2019) 2483–2499.
[30] F. Maciá-Pérez, J. V. Berna-Martinez, A. F. Oliva, M. A. A. Ortega, Algorithm for the detection of outliers based on the theory of rough sets, Decision support systems 75 (2015) 63–75.
[31] Y. Wang, Y. Li, Outlier detection based on weighted neighbourhood information network for mixed-valued datasets, Information Sciences 564 (2021) 396–415.
[32] Z. Yuan, H. M. Chen, T. R. Li, X. Y. Zhang, B. B. Sang, Multigranulation relative entropy-based mixed attribute outlier detection in neighborhood systems, IEEE Transactions on Systems, Man, and Cybernetics: Systems 52 (8) (2021) 5175–5187.
[33] Z. Yuan, H. Chen, T. Li, B. Sang, S. Wang, Outlier detection based on fuzzy rough granules in mixed attribute data, IEEE Transactions on Cybernetics 52 (8) (2021) 8399–8412.
[34] Z. Yuan, B. Y. Chen, J. Liu, H. M. Chen, D. Z. Peng, P. L. Li, Anomaly detection based on weighted fuzzy-rough density, Applied Soft Computing 134 (2023) 109995.
[35] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, International Journal of General System 17 (2-3) (1990) 191–209.
[36] 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.
[37] E. M. Knox, R. T. Ng, Algorithms for mining distance-based outliers in large datasets, in: Proceedings of the international conference on very large databases, Citeseer, 1998, pp. 392–403.
[38] 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.
[39] K. Zhang, M. Hutter, H. D. 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.
[40] 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.
[41] H. Xu, Y. Wang, Y. Wang, Z. Wu, Mix: A joint learning framework for detecting both clustered and scattered outliers in mixed-type data, in: 2019 IEEE International Conference on Data Mining (ICDM), IEEE, 2019, pp. 1408–1413.
[42] V. Hautamaki, I. Karkkainen, P. Franti, Outlier detection using k-nearest neighbour graph, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., Vol. 3, IEEE, 2004, pp. 430–433.

[43] C. You, D. P. Robinson, R. Vidal, Provable self-representation based outlier detection in a union of subspaces, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 3395–3404.
[44] X. Li, J. Lv, Z. Yi, Outlier detection using structural scores in a high-dimensional space, IEEE transactions on cybernetics 50 (5) (2018) 2302–2310.
[45] Z. Yuan, H. M. Chen, C. Luo, D. Z. Peng, Mfgad: Multi-fuzzy granules anomaly detection, Information Fusion 95 (2023) 17–25.

Declaration of interests

☒ 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: