Kernelized Fuzzy-Rough Anomaly Detection

Yan Wu, Sihan Wang, Hongmei Chen, Dezhong Peng, and Zhong Yuan

***Abstract*—Anomaly detection is a significant area of discovering knowledge that has shown success in the areas of fraud detection, cyber security, and medical diagnostics. The kernelized fuzzy-rough set is a key extension model in rough computing. It inherits the advantages of kernel function and fuzzy-rough set and can handle uncertain information in data more effectively. However, existing kernelized fuzzy-rough set models construct upper and lower approximation sets mainly from decision attribute data which are not available to unlabeled data. In addition, to the best of our knowledge, studies related to the use of kernelized fuzzy upper and lower approximation sets for constructing effective anomaly detection models have not yet been reported. Based on these observations, this paper constructs a kernelized fuzzy-rough set model for unlabeled data and proposes a kernelized fuzzy-rough anomaly detection (KFRAD) method. Specifically, we first optimize the kernel function to compute the fuzzy relation matrix of attribute subsets to construct the fuzzy approximation space. Then, the definition of kernelized fuzzy approximation accuracy for unlabeled data is given. Then, the granule anomaly extent is determined based on the kernelized fuzzy approximation accuracy. Finally, the granule anomaly extents and the corresponding weight values are integrated to compute the kernelized fuzzy-rough anomaly scores of the objects. On the basis of the above detection ideas, we design the KFRAD algorithm and experimentally compare it with mainstream detection algorithms. The analysis results show that the proposed algorithm has better performance. The code is publicly available online at https://github.com/wwwuyan/KFRAD.**

***Index Terms*—Granular computing, anomaly detection, kernel function, fuzzy-rough sets, unlabeled data**

I. Introduction

ANOMALY detection is a key technique in data analytics aimed at identifying anomalous objects in a dataset. Presently, anomaly detection methods are extensively employed in finance, cyber security, production quality inspection, and medical diagnosis. Anomaly detection techniques make assumptions about anomalous and normal objects in different ways and are typically classified into statistical [1], distance-based [2], density-based [3], depth-based [4], and cluster-based [5] methods. However, all these anomaly detection algorithms have certain limitations in practical applications. Statistical methods can only target a single attribute and are not applicable to multidimensional datasets. Depth-based algorithms are suitable for 2D or 3D spatial data and are less efficient in detecting high-dimensional data with mixed attributes. Distance-based algorithms suffer from sparsity problems when dealing with high-dimensional data and are very sensitive to parameter selection. Density-based algorithms have the same problem of parameter selection sensitivity. Furthermore, most distance-based and density-based algorithms are not suitable for dealing with categorical or mixed attribute data because they use Euclidean distances to construct detection models.

Rough set theory (RST) is a mathematical model based on set theory, which can be used to deal with uncertainty problems such as inconsistency and incompleteness. It has been widely applied in the fields of pattern recognition, machine learning, and data mining. For the past few years, in order to compensate for the drawback that Euclidean distance-based methods are not applicable to categorical (nominal) attribute data, rough anomaly detection techniques have also been proposed and refined over time [6]–[8]. However, these methods are only applicable to categorical data, which need to be discretized first when dealing with numerical data. As a result, this may cause a large amount of information loss and lead to distortion of the processing results.

In order to compensate for the deficiencies that exist in rough computing, the fuzzy-rough set (FRS) was introduced by Dubois and Prade [9]. Since FRS can transform numerical attributes into fuzzy values to represent the similarity between samples, it can deal with numerical attribute data directly without discretization, maintaining more authentic information. The kernel function is a concept of great importance in machine learning and data mining. The kernel method maps low-dimensional linearly indivisible data into a high-dimensional space through nonlinear mapping to achieve linear divisibility and converts the inner product operation in high-dimensional space into a kernel function computational problem in low-dimensional space. This cleverly solves the problem that it is difficult to compute in high-dimensional space and difficult to classify in low-dimensional space. In terms of how to generate effective fuzzy similarity relations, Hu et al. [10], [11] used for the first time a Gaussian kernel function satisfying reflexive, symmetric, and T_{cos}-transitive to compute fuzzy equivalence relations. Further, the kernelized fuzzy upper and lower approximation for the decision granules is defined, and thus a supervised feature selection method based on the kernelized fuzzy-rough set (KFRS) is proposed. So far, the theory of KFRS has been continuously developed and improved, and become an important research direction in machine learning and data mining. However, the existing KFRSs mainly study the approximation of decision granules, which cannot be applied to unlabeled data.

Based on the above discussion, for the problem that existing KFRSs are not applicable to unlabeled data, in this paper, we

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

Y. Wu, S. H. Wang, D. Z. Peng, and Z. Yuan are with the College of Computer Science, Sichuan University, Chengdu 610065, China (E-mail: wu_yan@stu.scu.edu.cn, wangsihan0713@foxmail.com, pengdz@scu.edu.cn, yuanzhong@scu.edu.cn

H. M. Chen is with the School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China (E-mail: hmchen@swjtu.edu.cn).

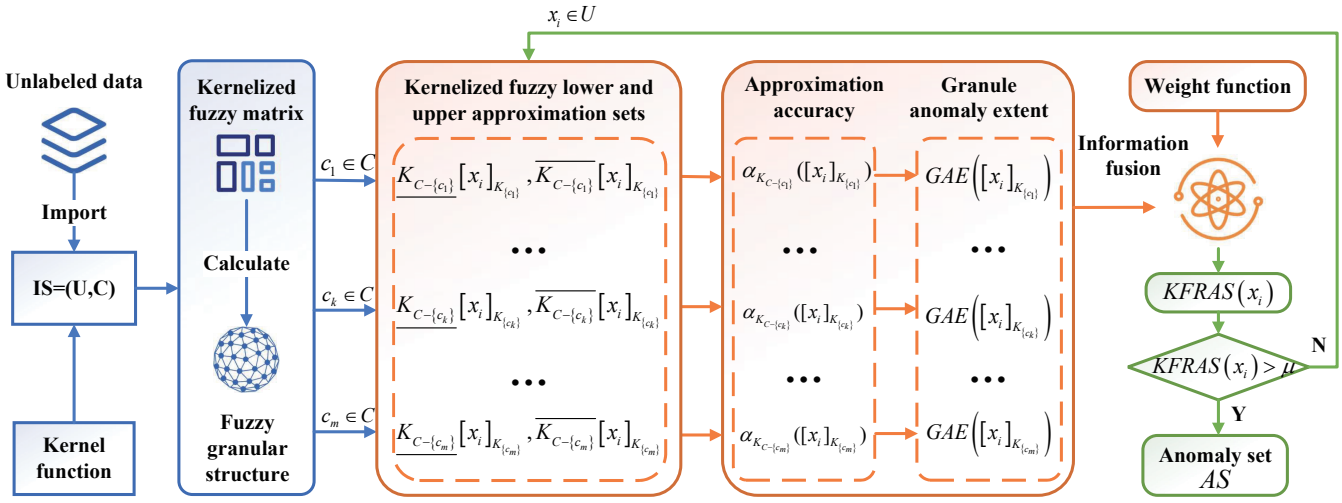


Fig. 1: The framework of the KFRAD method

first construct the KFRS model oriented to unlabeled data. On this basis, the kernelized fuzzy-rough anomaly detection (KFRAD) method is presented. We use the kernel function to generate fuzzy information granules from the data and construct the KFRS for unlabeled data by approximating the fuzzy information granules. Then, the notion of kernelized fuzzy approximation accuracy is given. On this basis, the granule anomaly extent (GAE) is defined to indicate the extent of abnormality of fuzzy information granules. Finally, the kernelized fuzzy-rough anomaly score (KFRAS) is computed by fusing GAEs and the corresponding weights to characterize the anomaly degree of the sample, and then the KFRAD algorithm is designed. The proposed detection algorithm combines the respective advantages of kernel functions and FRSs, thus it is suitable for unsupervised anomaly identification with uncertainty and complexity. We experimentally compare the KFRAD algorithm with mainstream anomaly detection methods in related UCI data experiments. The results show that the KFRAD algorithm is effective and further improves the performance of anomaly detection. The framework of the KFRAD method is illustrated in Fig. 1. Overall, our main contributions to this work are as follows.

1) The Gaussian kernel is used to compute fuzzy similarity relations.
2) A KFRS model for unlabeled data is proposed.
3) An anomaly detection algorithm is constructed based on the KFRS model for unlabeled data.
4) Experimental results demonstrate that the method has better performance and adaptability.

The paper is organized as follows. Section II briefly introduces the related work on rough anomaly detection, extended rough anomaly detection, and KFRSs. Section III reviews some preliminary knowledge about KFRSs. Section IV proposes a KFRS model for unlabeled data. Section V introduces a new kernelized fuzzy-rough anomaly detection model and the related algorithm. Section VI compares the proposed algorithm with some algorithms experimentally. Finally, we summarize the whole paper.

II. Related Work

This section discusses some rough anomaly detection methods and extended rough anomaly detection methods, as well as related research on KFRSs.

A. Rough anomaly detection methods

In order for the drawbacks of Euclidean distance-based methods to be compensated, researchers have presented various rough anomaly detection methods [6], [7], [12]. For instance, the boundary-based outlier detection method by utilizing the RS framework to detect outliers was presented by Jiang et al [12]. Chen et al. [7] introduced an outlier detection method based on granular computing (GrC). Jiang et al. [8] proposed a definition of outlier detection based on rough sequences and discussed an outlier detection method based on the distance metric. In addition, an outlier detection method based on information entropy (IE) was presented in [13], which draws on the classical information entropy model. An outlier approximation approach based on the anomaly detection and analysis system was introduced in [14] by utilizing the concept of RS, which can mine outliers in a subset of key attributes by defining the similarity of peripheral partitions. A hybrid anomaly detection method that combines the perspectives of boundary-based and distance-based anomaly detection was proposed in [15]. Albanese et al. [16] presented a rough outlier set extraction method that finds the highest outliers in unlabeled spatio-temporal datasets to deal with the problem of outlier detection in spatio-temporal data. A novel anomaly detection combining GrC and RST based on classical approximation accuracy was presented in [17]. Maciá-Pérez et al. [18] suggested a powerful method based on RST that can be used to find outliers in plenty of data. However, the above method relies on equivalence relations to partition the set of objects and generate disjoint object equivalence classes. Therefore, it is solely suitable for data with nominal (categorical) attributes.

B. Extended rough anomaly detection methods

With the deepening research on RST, many academics have extended rough set models to address issues like the rough set model's inapplicability to data with numerical attributes. Common extended rough set models include the FRS model [9], fuzzy neighborhood rough set model [19], KFRS model [11], and so on. These extended RS models have been successfully applied in the direction of anomaly detection. For instance, Chen et al. [20] introduced the neighborhood model as a uniform framework to comprehend and apply anomaly detection methods. The hybrid data-driven neighborhood entropy-based anomaly detection approach for three data types in [21] extended the standard distance-based and rough set-based anomaly detection methods. [22] presented a weighted neighborhood information network that represented mixed-value attribute datasets by considering object similarities and neighborhood relations. In [23], a fuzzy information entropy-based anomaly detection method for mixed-feature anomalies was developed. A fuzzy-rough processing framework to handle outlier detection in mixed data was introduced in [24]. On the basis of the multi-fuzzy granules computation model, Yuan et al. [25] constructed forward and reverse multi-fuzzy granules to define the anomaly scores for anomaly detection. An anomaly detection method based on fuzzy-rough density was presented in [26] considering both sample density and uncertainty information. Nowadays, extended rough anomaly detection approaches have received a lot of attention and achieved better detection performance. However, most of the methods for fuzzy-rough computing use the strategy of intersection to construct the fuzzy granular structure, which may lead to the inability to accurately characterize the similarity between the samples in high-dimensional space.

C. Kernelized fuzzy-rough sets

The FRS which has been extensively investigated and used in many fields is a major extension of RS. However, how to select fuzzy relations to improve the performance of the fuzzy-rough model is still a problem to be discussed. The choice of fuzzy relations will directly impact how well the model performs, so it's critical to have an organized and efficient process for determining fuzzy relations. Moser [27] proved that any kernel is at least T_{cos}-transitive when it satisfies reflexive and symmetric. Relations calculated via such kernel functions are fuzzy T-equivalence relations that can be used to granulate the domain of the argument to produce a class of fuzzy information granules. Subsequently, these fuzzy information granules can be used to approximate various subsets of the universe. By constructing different kernelized granular spaces, FRSs on different kernel functions can be formed, i.e., KFRSs [10], [11].

KFRS can compensate for the inability of intersection operations to accurately describe the similarity between samples. It has been effectively used in many knowledge discovery missions as a member of the kernel machine family. For example, Hu et al. [11] discussed a feature evaluation system based on KFRS, which provided a new perspective for understanding and extending rough sets and proposed a generalized feature evaluation function and an attribute approximation algorithm. Further, a KFRS model based on Gaussian kernel approximation was presented in [10], demonstrating the application of Gaussian kernel approximation in feature evaluation and feature approximation. Chen et al. [28] used the FRS technique to compute the extent of affiliation, which enabled the hard interval support vector machine to be combined with the FRS and presented a FRS based on the fuzzy transfer kernel. Qu et al. [29] made improvements on the basis of existing fuzzy-rough nearest neighbor classifiers and introduced a kernelized fuzzy-rough nearest neighbor classification algorithm. Li et al. [30] fused the kernelized information from feature space and label space, and combined FRSs with multi-kernel learning to implement a multilabel learning feature selection method based on KFRS. Although the research related to KFRS has involved feature selection, dimensionality reduction, clustering, and other tasks and achieved better results, the research on anomaly detection with KFRS has to be further explored.

III. Preliminaries

In this section, we go over some definitions and notations about KFRSs.

In granular computing theory, data are imported into an Information System (IS), which is formally defined as $IS = (U, A)$, where U denotes a finite set of non-empty samples and A denotes a finite set of non-empty attributes. When $A = C \cup D$ and $C \cap D = \emptyset$, the data table is called a Decision System (DS), where C denotes the set of conditional attributes and D denotes the set of decision attributes.

Definition 1: [11] Given a non-empty finite set U, a real-valued function $k : U \times U \to R$ is said to be kernel if it satisfies semipositivity and symmetry for any $x, y \in U$.

Theorem 1: [11] Any kernel $k : U \times U \to [0, 1]$ with $k(x, x) = 1$ satisfies (at least) T_{cos}-transitive, where

$$T_{cos}(a, b) = \max\left\{ab - \sqrt{1-a^2}\sqrt{1-b^2}, 0\right\}. \quad (1)$$

Since some kernel functions are simultaneously reflexive, symmetric, and T_{cos}-transitive, fuzzy T-equivalence relations are satisfied by the relations computed with these kernel functions. Gaussian kernel, Exponential kernel, and Rational quadratic kernel [11], etc., are common kernel functions that meet the aforementioned qualities and can be used to create KFRSs by swapping out fuzzy relations in FRSs. We use K_B to denote the fuzzy relation computed with the kernel function with respect to the attribute subset $B \subseteq A$.

Definition 2: Let $U = \{u_1, u_2, \cdots, u_n\}$ be a non-empty finite set of objects, and the kernelized fuzzy relation K_B on U refers to a fuzzy set on $U \times U$.

The kernelized fuzzy relation K_B on U can be expressed by a kernelized fuzzy matrix, i.e., $M_{K_B} = (k_{ij}^B)_{n\times n}$, where $k_{ij}^B = K_B(u_i, u_j)$, and a fuzzy set is represented by each row vector. The set of all fuzzy relations on U is denoted as $F(U)$.

Theorem 2: Let K_B be a kernelized fuzzy relation on U. For any $x, y \in U$, The properties listed below are met.

1) $K_B(x, y) \in (0, 1]$;
2) $K_B(x, x) = 1$;
3) $K_B(x, y) = K_B(y, x)$.

In addition, Moser showed that K_B also satisfies T_{cos} transitive [27]. Obviously, kernelized fuzzy relations are a special class of fuzzy sets with the properties of fuzzy sets, and also exist operations such as merge, intersection, and complementation. In order to deal with uncertainty and ambiguity information more accurately, the fuzzy set operations are extended to triangular-norm (t-norm), triangular-conorm (s-norm or t-conorm), and negator [11]. In the following, T_{cos}, S_{cos}, and N are used to denote the t-norm, the s-norm, and the negator, respectively.

Definition 3: Given a non-empty finite set of objects U and kernelized fuzzy relations K_B. Let $X \in F(U)$, the kernelized fuzzy upper and lower approximations of X with respect to K_B are defined as

$$\overline{K_B}X(x) = \sup_{y\in U} T_{cos}(K_B(x,y), X(y)), \tag{2}$$

$$\underline{K_B}X(x) = \inf_{y\in U} S_{cos}(N(K_B(x,y)), X(y)), \tag{3}$$

where,

$T_{cos}(a,b) = \max\{ab - \sqrt{1-a^2}\sqrt{1-b^2}, 0\}$,

$S_{cos}(a,b) = \min\{a+b-ab - \sqrt{2a-a^2}\sqrt{2b-b^2}, 1\}$.

Since the kernelized fuzzy relations K_B is a T-equivalence relation, $\overline{K_B}$ and $\underline{K_B}$ satisfy the properties mentioned in [31].

Although the model of KFRS has been discussed more comprehensively [29], [32], to the best of our knowledge, the research on the KFRS model is almost exclusively focused on the approximation of the decision information granules [11]. Specifically speaking, the $X(y)$ here is always 1 or 0, i.e., belonging to the same class as x or not. This research method can only be oriented to decision attribute data and cannot be applied to unlabeled data. In the next section, we consider constructing KFRS models oriented to unlabeled data by extracting fuzzy relations directly from the data using kernel functions.

IV. Kernelized Fuzzy-Rough Set Model for Unlabeled Data

Most of the existing KFRS models are oriented to labeled data, which is very limited in dealing with unlabeled data. In this section, a KFRS model for unlabeled data is proposed based on the existing results of KFRS models.

Let $U = \{x_1, x_2, \ldots, x_n\}$, $C = \{c_1, c_2, \ldots, c_m\}$, for any $B \subseteq C$, let $B = \{c_{t_1}, c_{t_2}, \ldots, c_{t_h}\} (h \in [1, m])$, one can derive the kernelized fuzzy relation K_B for B on U, and $M_{K_B} = \left(k_{ij}^B\right)_{n\times n}$ denotes the kernelized fuzzy matrix of K_B, where $k_{ij}^B = K_B(x_i, x_j)$. In this study, the Gaussian kernel function is employed to calculate the fuzzy similarity relations among objects, and then we construct the KFRS model for unlabeled data and apply it to anomaly detection.

Definition 4: Given a finite set of non-empty samples U and a set of conditional attributes C, for any $B \subseteq C$, the fuzzy information granular structure $G_U(K_B)$ of U induced by the kernelized fuzzy relation K_B is defined as

$$G_U(K_B) = \left\{[x_1]_{K_B}, [x_2]_{K_B}, \ldots, [x_n]_{K_B}\right\}. \tag{4}$$

A kernelized fuzzy relation can derive a fuzzy granular structure, where $[x_i]_{K_B} = \left(k_{i1}^B / x_1\right) + \left(k_{i2}^B / x_2\right) + \cdots + \left(k_{in}^B / x_n\right) = \left(k_{i1}^B, k_{i2}^B, \ldots, k_{in}^B\right)$ is known as the fuzzy information granules produced by the kernelized fuzzy relation K_B.

Clearly, $[x_i]_{K_B}$ is a fuzzy set on K_B. where $[x_i]_{K_B}(x_j) = K_B(x_i, x_j) = k_{ij}^B$. If $K_B(x_i, x_j) = 0$, it shows that x_j must not belong to $[x_i]_{K_B}$, and if $K_B(x_i, x_j) = 1$, it shows that x_j must belong to $[x_i]_{K_B}$. The formula for the fuzzy information granules base is $\left|[x_i]_{K_B}\right| = \sum_{j=1}^{n} K_B(x_i, x_j)$. We can find $1 \le \left|[x_i]_{K_B}\right| \le n$.

Definition 5: Given a non-empty finite set of objects U and a set of conditional attributes C, for any $P, B \subset C$, $G_U(K_B) = \left\{[x_1]_{K_B}, [x_2]_{K_B}, \ldots, [x_n]_{K_B}\right\}$. For any $[x_i]_{K_B} \in G_U(K_B)$, the kernelized fuzzy upper and lower approximations of $[x_i]_{K_B}$ with respect to K_P are a pair of fuzzy sets on U, which are computed as follows.

$$\overline{K_P}[x_i]_{K_B}(x) = \sup_{y\in U} T_{cos}(K_P(x,y), [x_i]_{K_B}(y)), \tag{5}$$

$$\underline{K_P}[x_i]_{K_B}(x) = \inf_{y\in U} S_{cos}\left(N(K_P(x,y)), [x_i]_{K_B}(y)\right). \tag{6}$$

It is easy to find that $\overline{K_P}[x_i]_{K_B}$ and $\underline{K_P}[x_i]_{K_B}$ approximate different goals from existing kernelized fuzzy upper and lower approximations. The existing kernelized fuzzy upper and lower approximations approximate X which is often a subset of objects induced by a decision attribute [11], whereas the kernelized fuzzy upper and lower approximations for unlabeled data directly approximate the fuzzy information granules generated from the conditional attribute data through the computation of Gaussian kernel functions.

Theorem 3: Given a non-empty finite set of objects U, a set of conditional attributes C, and the kernelized fuzzy relations K_P, K_B, where $\forall P, B \subset C$, for any $[x_i]_{K_B} \in G_U(K_B)$, $\overline{K_P}[x_i]_{K_B}$ and $\underline{K_P}[x_i]_{K_B}$ satisfy the following properties.

1) $\underline{K_P}[x_i]_{K_B} \subseteq [x_i]_{K_B} \subseteq \overline{K_P}[x_i]_{K_B}$;
2) $\underline{K_P}(\underline{K_P}[x_i]_{K_B}) = \underline{K_P}[x_i]_{K_B}, \overline{K_P}(\overline{K_P}[x_i]_{K_B}) = \overline{K_P}[x_i]_{K_B}$;
3) For any $M \subseteq C$, if $K_M \subseteq K_P$, then $\underline{K_P}[x_i]_{K_B} \subseteq \underline{K_M}[x_i]_{K_B} \subseteq [x_i]_{K_B} \subseteq \overline{K_M}[x_i]_{K_B} \subseteq \overline{K_P}[x_i]_{K_B}$.

We replace the decision information granules of the samples with the kernelized fuzzy information granules $[x_i]_{K_B}$ extracted from the conditional attribute data and construct the corresponding upper and lower approximations. This enables efficient processing of unlabeled data and provides a theoretical basis for the construction of the kernelized fuzzy-rough anomaly detection model later.

V. Kernelized Fuzzy-Rough Anomaly Detection

Based on the KFRS for unlabeled data, this section defines a new anomaly score calculation and further constructs the kernelized fuzzy-rough anomaly detection method, which mainly includes two parts: detection model and algorithm.

A. Detection model

This subsection demonstrates the anomaly detection approach based on the KFRS, which mainly includes normalization preprocessing, calculating fuzzy relation matrix for attribute subsets, calculating approximation accuracies, calculating anomaly scores, and anomaly discrimination.

There are frequently disparities in the magnitude and dimension of data in the data analysis of information systems. In order to avoid the impact of such differences on data processing, the original data should be normalized [21]. Common normalization methods include min-max normalization, decimal place normalization, and Z-Score normalization. In this study, the min-max normalization method is used to preprocess the data so that all the numerical attributes of the data range between $[0, 1]$. The relevant formula is as follows.

$$f(c_t(x_i)) = \frac{c_t(x_i) - \min_{c_t}}{\max_{c_t} - \min_{c_t}}, \tag{7}$$

where $c_t(x_i)$ denotes the value of the attribute c_t on x_i, and $\max_{c_t}$ and $\min_{c_t}$ are the maximum and minimum values in the domain U with respect to c_t, respectively.

Gaussian kernel functions are a widely used class of kernel functions in fields such as support vector machines and neural networks. Nonlinear problems are mapped to higher dimensional spaces by Gaussian kernels, which often lead to better performance and computational efficiency. For a subset of attributes $B \subseteq C$, in order to compute the kernelized fuzzy relation matrix, we use the Gaussian kernel function to compute the fuzzy similarity between x_i and x_j with respect to B, defined as follows.

$$k_{ij}^B = \exp(-\frac{\|B(x_i) - B(x_j)\|^2}{\delta}), \tag{8}$$

where δ is the Gaussian kernel parameter and $\|B(x_i) - B(x_j)\|$ is the Euclidean distance between two objects with respect to a subset of attributes B.

Let $B = \{b_1, b_2, \ldots, b_h\}\,(h \in [1, m])$, the Euclidean distance between x_i and x_j with respect to the subset of attributes B is computed as follows.

$$\|B(x_i) - B(x_j)\| = \sqrt{\textstyle\sum_{l=1}^{h} (b_l(x_i) - b_l(x_j))^2}. \tag{9}$$

According to $M_{K_B} = (k_{ij}^B)_{n\times n}$, the fuzzy relation matrix can be derived. Obviously, when $h = 1$, the single-attribute fuzzy relation matrix is computed, and when $h > 1$, the multi-attribute fuzzy relation matrix is computed.

Next, the kernelized fuzzy-rough anomaly detection method is constructed on the basis of the KFRS model for unlabeled data. As shown in Section IV, the fuzzy information granules $[x_i]_{K_B}$ can be constructed for $x_i \in U$ and kernelized fuzzy similarity relation K_B. In the process of doing anomaly detection, we first calculate the anomaly extent of $[x_i]_{K_B}$, A higher granule anomaly extent indicates that the object is more distinct from other objects on the subset of attributes B. A weighted average of several granule anomaly extents will calculate an anomaly score, which can indicate the overall level of abnormality of the object. Thus in general, the probability that an object is an anomaly is positively correlated with its anomaly score.

In order to compute the anomaly extent of fuzzy information granules, the approximation accuracy of KFRS is defined. The approximation accuracy of KFRS depends on the selection and implementation of the upper and lower approximation algorithms. Based on the upper and lower approximation formulas of KFRS oriented to unlabeled data, we define the kernelized approximation accuracy as follows.

Definition 6: For any $B \subset C$, where $|C - B| \geq 2$, let $G_U(K_B) = \left\{[x_1]_{K_B}, [x_2]_{K_B}, \ldots, [x_n]_{K_B}\right\}$, for any $[x_i]_{K_B} \in G_U(K_B)$ and $P \subseteq (C - B)$, the approximation accuracy of $[x_i]_{K_B}$ with regard to K_P is defined as

$$\alpha_{K_P}\left([x_i]_{K_B}\right) = \frac{\left|\underline{K_P}\,[x_i]_{K_B}\right|}{\left|\overline{K_P}\,[x_i]_{K_B}\right|}, \tag{10}$$

Obviously, it can be seen that $0 \leq \alpha_{K_P}\left([x_i]_{K_B}\right) \leq 1$.

Here, we characterize the anomaly extent of fuzzy information granules in terms of approximation accuracy. For a particular x_i, we compute the approximation accuracy of $[x_i]_{K_B}$ with respect to a set of fuzzy similarity relations. When the approximation accuracy is constantly poor, the anomaly scores of x_i will also be higher, and it is more likely to be regarded as an anomaly object. For any $x_i \in U$, the Boolean attribute of x_i is provided by the majority of existing anomaly detection algorithms, i.e., whether it is an anomaly or not, but in practice, it is often more meaningful to compute the anomaly score of x_i. Therefore, two concepts are presented next in this paper, (1) GAE and (2) KFRAS. GAE is used to quantify the anomaly extent of fuzzy information granules, and KFRAS is used to denote the anomaly score of an object.

Definition 7: Let $P = C - B = \{p_1, p_2, \ldots, p_q\}$, for any $x_i \in U$, the $\mathrm{GAE}([x_i]_{K_B})$ is defined as

$$\mathrm{GAE}\left([x_i]_{K_B}\right) = 1 - \frac{\left|[x_i]_{K_B}\right| \cdot \left(\alpha_{K_P}\left([x_i]_{K_B}\right)\right)}{|U|}. \tag{11}$$

Since $\alpha_{K_P}\left([x_i]_{K_B}\right)$ can be used to measure the uncertainty in $[x_i]_{K_B}$, we use $\alpha_{K_P}\left([x_i]_{K_B}\right)$ to compute $\mathrm{GAE}\left([x_i]_{K_B}\right)$ and use it to measure the anomaly extent of $[x_i]_{K_B}$.

Since a fuzzy similarity relation can be computed for each subset of attributes of C, to compute the approximation accuracy of all fuzzy similarity relations, we would need $2^{|C|}$ fuzzy similarity relations. Accordingly, the time complexity would be exponential in relation to $|C|$. However, it is clearly undesirable and impractical considering the operational efficiency of the program. Therefore, we only compute the approximation accuracy of K_P.

Definition 8: For any $x_i \in U$, the $\mathrm{KFRAS}(\mathrm{x_i})$ is defined as

$$\mathrm{KFRAS}\,(x_i) = \frac{\sum_{t=1}^{m}\left(\mathrm{GAE}\left([x_i]_{K_{c_t}}\right) \cdot W_{c_t}\,(x_i)\right)}{|C|}, \tag{12}$$

where $W_{c_t} : U \to [0, 1]$ is the weight function satisfying $W_{c_t}(x_i) = 1 - \sqrt[3]{\left|[x_i]_{K_{c_t}}\right| / |U|}$.

For each $B \subset C$, a fuzzy similarity matrix K_B can be determined, and a GAE will be calculated for each K_B, so the total number of GAEs will reach $2^{|C|}$. However, due to its exponentially increasing time complexity, it is impractical to compute the anomaly scores of all GAEs. In this paper, we use only the $|C|$ GAEs determined by $c_t \in C$ to compute $\mathrm{KFRAS}(x_i)$ in order to simplify complexity.

The computation of the weight function W_{c_t} satisfies the idea that anomaly detection tends to favor a small number of objects in a dataset and that a few objects are more likely to be

anomalies than the majority of objects. For any $x_i \in U, c_t \in C$, if the cardinality of $[x_i]_{K_{c_t}}$ is smaller than that of the other fuzzy information granules, x_i can be regarded as a small number of objects on the attribute c_t and a higher weight can be assigned to the fuzzy information granule. From Formula 12, if the weight is heavy, the GAE will play an important role in the calculation of KFRAS. When the GAEs of an object always stay high, the KFRAS obtained from their weighted summation will also be high. It means that the object is more likely to be an anomaly.

Given an anomaly threshold μ, for any $x_i \in U$, we say that x_i is an anomaly object in U if $\mathrm{KFRAS}(x_i) > \mu$. In order to determine the threshold μ, the users should first input an empirical value (denoted by symbol eon) based on their expected number of anomalies. For a given dataset, the determination of eon needs numerous trials and the eon of various datasets are always different. For the anomaly score, we first arrange it in descending order to get a sequence $s_1^{'}, s_2^{'}, \ldots, s_n^{'}$, and then set the range of μ to $[s_{eon+1}^{'}, s_{eon}^{'})$. As a result, When the threshold μ is too low, more normal objects may be considered to be anomalies. When the threshold is too high, more anomaly objects will likely be difficult to detect.

B. Detection algorithm

The kernelized fuzzy-rough anomaly detection approach was introduced in the preceding section, and in this section, we will design the corresponding KFRAD algorithm and analyze its complexity.

In Algorithm 1, three cycles are used to obtain the set of anomalies. The first cycle computes the kernelized fuzzy relation matrix for single-attribute subsets. The second to third nested cycle first computes the kernelized fuzzy relation matrix for multi-attribute subsets and then calculates approximation accuracy for fuzzy information granules, on the basis of which it computes the GAEs and related weights. Furthermore, it computes the KFRAS for each object and finally decides whether to put it into the set of anomalies or not through the anomaly threshold. The complexity of the algorithm is analyzed below. Steps 2-4 are iterated $|C|$ times, Step 3 is iterated $|U|$ times, Steps 5-16 are iterated $|U|$ times, Steps 6-11 are iterated $|C|$ times, and Step 7 is iterated $|U|$ times. It can be seen that the overall number of KFRAD algorithm cycles is $|C| \times |U| + |U| \times |C| \times |U|$. Consequently, the time complexity of KFRAD is $O(|C||U|^2)$.

Algorithm 1: KFRAD Algorithm

Input: $IS = (U, C), \mu, \delta$.
Output: Anomaly Set(AS).

```
AS ← ∅;
for t ← 1 to m do
    Calculate M_{K_{c_t}};
end
for i ← 1 to n do
    for t ← 1 to m do
        Calculate M_{K_{C-{c_t}}}; //Let B = c_t.
        Calculate α_{C-{c_t}}([x_i]_{K_{c_t}}); //Let P = C − B.
        Calculate GAE([x_i]_{K_{c_t}});
        Calculate W_{c_t}(x_i);
    end
    Calculate KFRAS(x_i);
    if KFRAS(x_i) > μ then
        AS ← AS ∪ {x_i};
    end
end
return AS.
```

VI. Experiments and Analyzes

In this section, we conduct an extensive comparative study of KFRAD with nine mainstream algorithms to demonstrate its better performance. Meanwhile, we analyze the effect of attribute noise on AUC, the sensitivity of the parameters, and the statistical significance of the results. Before the experiment, we first do some preparatory work, such as dataset preparation, experimental parameter setting, and evaluation index setting.

A. Experimental preparation

The dataset we have chosen comes from publicly available web pages [1,2], which are often used in evaluation efforts of anomaly detection methods. In the experiment, we selected 16 datasets from them. As can be seen from Table I, the data sample sizes range from 94 to 4429, while the data dimensions range from 4 to 69.

On 16 datasets, we experimentally compared the KFRAD algorithm with 9 algorithms. The experimental algorithms and corresponding time complexities are illustrated in Table II, where z, ϕ, r, E, T denote the testing data size, the subsampling size, the number of trees, the time complexity of deep autoencoder and the number of iterations. We set the optimal parameters of the algorithms according to the strategy in the corresponding paper. For VarE, The optimal parameters are obtained in $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3\}$. NOF, ECOD, and ROD are parameter-free. The parameters of DCROD and NC are obtained in $[1, 60]$ with a step size of 1. The base estimate of IForest is set to 100. The parameter of MIX is obtained in $[0, 9]$ with a step size of 1, and the parameter of SRO is obtained in $[1, 20]$ with a step size of 1. Due to the time overhead and equipment limitations, we experimented with some smaller datasets beforehand and found that the range of values of the optimal parameters of KFRAD in this paper is generally in $[0.02, 0.8]$. As a result, we set the parameter of KFRAD in the range of $[0.02, 0.8]$ with a step size of 0.02. During preprocessing, all the different nominal attribute data are represented as different integers. Additionally, all attribute values are normalized to $[0, 1]$ by the min-max normalization method.

In this paper, to better evaluate the performance of the model on an unbalanced dataset, where the number of anomaly

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

TABLE I: The description of datasets

ID	Datasets	Abbreviation	Unlabeled attribute	Object	Anomaly
1	Audiology_variant1	Aud	69	226	53
2	Chess_nowin_145_variant1	Chess	36	1814	145
3	CreditA_plus_42_variant1	Cred	15	425	42
4	Diabetes_tested_positive_26_variant1	Diab	8	526	26
5	German_1_14_variant1	Germ	20	714	14
6	Heart_2_16_variant1	Heart	13	166	16
7	Hepatitis_2_9_variant1	Hepa	19	94	9
8	Horse_1_12_variant1	Horse	27	256	12
9	Iris_Irisvirginica_11_variant1	Iris	4	111	11
10	Monks_0_12_variant1	Monks	6	240	12
11	Mushroom_p_221_variant1	Mush	22	4429	221
12	Pima_TRUE_55_variant1	Pima	9	555	55
13	Tic-tac-toe_negative_12_variant1	Tic	9	638	12
14	Wisconsin_malignant_39_variant1	Wbc	9	483	39
15	Wdbc_M_39_variant1	Wdbc	31	396	39
16	Wine	Wine	13	129	10

TABLE II: The description of experimental algorithms

No.	Algorithms (Year)	Descriptions	Time complexities
1	VarE (2018) [33]	Outlier Detection Using Variance Structural Scores	$O(\|C\|\|U\|^2)$
2	NOF (2016) [34]	Natural Neighbor-based Outlier Detection	$O(\|U\|\log\|U\|)$
3	DCROD (2022) [35]	Directed density ratio Changing Rate-based Outlier Detection	$O(\|C\|\|U\|\log\|U\|)$
4	IForest (2008) [36]	Isolation Forest-based Anomaly Detection	$O(z\phi^2 + rz\phi)$
5	ECOD (2022) [1]	Empirical Cumulative distribution functions for Outlier Detection	$O(\|C\|\|U\|)$
6	MIX (2019) [37]	A Joint Learning-based Outlier Detection in MIXed-type data	$O(ET)$
7	NC (2016) [38]	Representation-based Boundary Point and Outlier Detection	$O(\|U\|^2)$
8	ROD (2020) [39]	Rotation-based Outlier Detection approach	$O(\|C\|^3\|U\|)$
9	SRO (2017) [40]	Self-Representation based Outlier detection	$O(\|U\|^2T)$
10	KFRAD (Ours)	Kernelized Fuzzy-Rough Anomaly Detection	$O(\|C\|\|U\|^2)$

objects is significantly smaller than the normal objects, we mainly use the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) to evaluate the performance of the anomaly detection algorithm [41], [42]. When comparing the performance of different anomaly detection algorithms using the ROC curve, the larger the area under the curve indicates the better performance. However, in practice, sometimes the ROC curve cannot clearly indicate which algorithm works better. We further use AUC to more accurately and quantitatively compare the performance of different algorithms. It is easy to find that the value of AUC is generally between $[0, 1]$, and the closer the AUC is to 1, the better the detection effect of the algorithm.

B. Experimental results

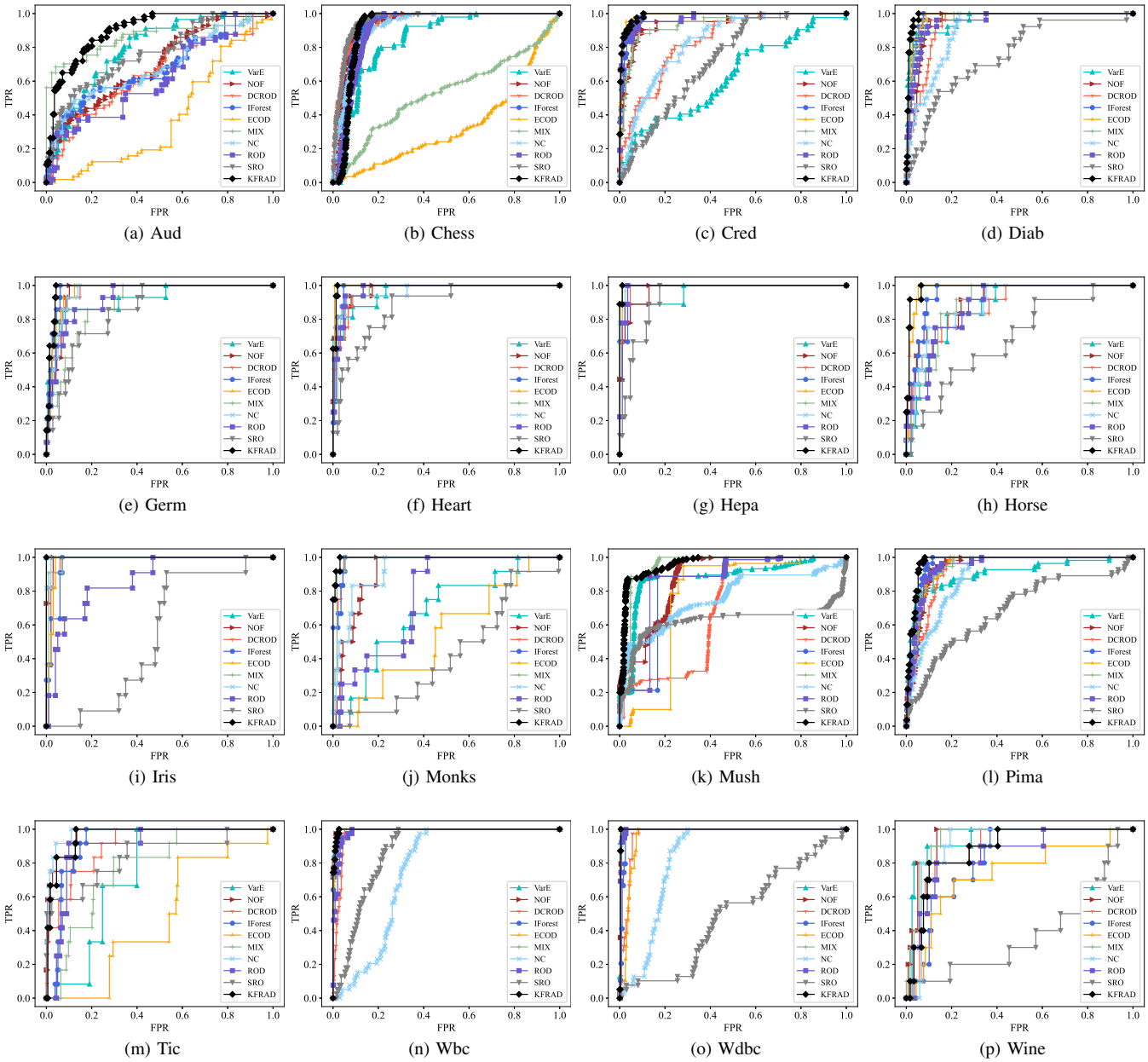
Firstly, we compare and analyze the results of the ROC curves. The ROC curves for each algorithm are depicted by different lines in Fig. 2, where the KFRAD algorithm is represented by a black line. As can be seen from the above description, the closer the ROC curve is to the upper left corner of the first quadrant, the better the results are indicated. On the Aud, Diab, Germ, Horse, Iris, Monks, and Pima datasets, it is easy to observe that the ROC curve of the KFRAD algorithm is closest to the upper left corner of the first quadrant. In particular, the curve on Aud is much higher than the curves of the other algorithms, indicating that the performance of KFRAD is much better on Aud than the other algorithms. In addition, the curves on Iris completely overlap with the coordinates, indicating that the performance of KFRAD is fully reflected on Iris. Also, the ROC curve of KFRAD is closer to the upper left corner of the first quadrant on the Cred, Heart, Hepa, and Wbc datasets. As a result, we may conclude that KFRAD performs better in the majority of circumstances.

However, on some datasets, there are cases where the ROC curves of KFRAD and the comparison algorithms overlap, e.g., in the Cred dataset, the ROC curves of KFRAD and ECOD overlap, which makes it difficult to compare their superiority intuitively. Therefore, we further calculated the AUC values of different algorithms on datasets for quantitative analysis.

Table III gives the comparison results of the AUC experiments, where the best AUC values are indicated in bold. The performance of KFRAD is analyzed using the Monks dataset as an example, the AUC of KFRAD is 0.996, which is much larger than VarE, ECOD, ROD, and SRO, and slightly larger than DCROD, IForest, and MIX. Statistically, KFRAD achieves the best results on 11 datasets, while the rest of the algorithms VarE, NOF, DCROD, IForest, ECOD, MIX, NC, ROD, and SRO achieve the best results on only 3, 2, 0, 0, 2, 2, 1, 0, and 1 datasets, respectively.

In addition, from Table III, it can be seen that for each dataset we ranked the AUC values in ascending order. When scores are equal, an average ranking of equal scores is calculated and this average ranking is applied to all algorithms with equal scores. Then the average ranking of each algorithm is calculated as 4.9, 6.4, 4.8, 6.4, 5.3, 6.5, 4.7, 4.7, 2.3, and 9.0 respectively. KFRAD also achieves the highest average ranking compared to the comparison algorithms.

The lowest AUC value calculated by KFRAD is 0.882 on



(a) Aud (b) Chess (c) Cred (d) Diab

(e) Germ (f) Heart (g) Hepa (h) Horse

(i) Iris (j) Monks (k) Mush (l) Pima

(m) Tic (n) Wbc (o) Wdbc (p) Wine

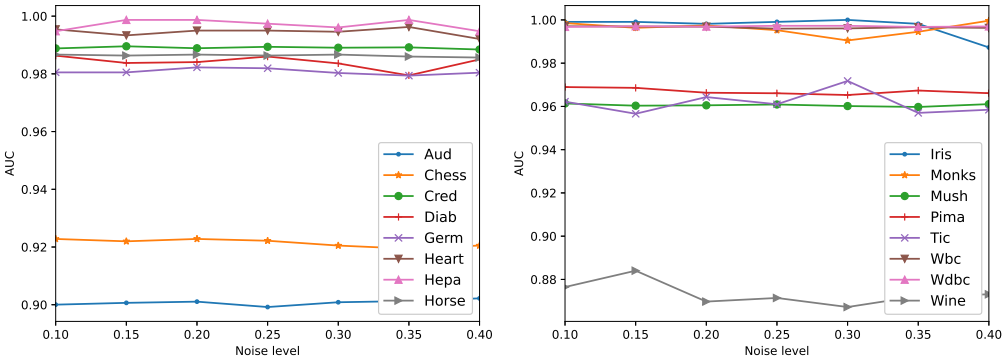
Fig. 2: Experimental comparison results on ROC

the Wine dataset. The reason it underperforms on Wine may be due to the fact that the number of samples and attributes in Wine is not very large. KFRAD is completely based on samples to construct fuzzy similarity relations, which has a large dependence on the size of samples and attributes, when the number of samples is small, KFRAD may be difficult to get the similarity relations comprehensively, which leads to a slightly poorer detection effect. Nevertheless, KFRAD performs well on the Iris dataset with similar dataset size, which may be related to the quality of the dataset.

Analyzed from a theoretical point of view, the kernel function allows data to be mapped from the original feature space to a high-dimensional feature space, thus making the original data easier to be linearly separated in the new feature space. For anomaly detection tasks, this means that with the kernel function, we can better capture the complex relations between anomalous and normal points, thus improving the accuracy of the detection. In the anomaly detection task, the objects may have diverse manifestations and the boundary among them may not be clear enough. And the FRS can help us to better deal with this uncertainty. The KFRS combines the advantages of both the two stages of feature extraction and anomaly detection of data respectively. KFRAD is built under the framework of the KFRS oriented to unlabeled data, which can extract kernelized fuzzy relation directly from the conditional attribute data through kernel calculation. It makes up for the defect that the existing KFRSs can only process decision attribute data through intersection operations. There-

TABLE III: Comparative results of AUC experiments

Datasets	VarE	NOF	DCROD	IForest	ECOD	MIX	NC	ROD	SRO	KFRAD
Aud	0.792 (8)	0.694 (6)	0.645 (3)	0.685 (4)	0.385 (1)	0.863 (9)	0.686 (5)	0.615 (2)	0.765 (7)	**0.900 (10)**
Chess	0.848 (3)	0.939 (8)	0.951 (9)	0.928 (7)	0.335 (1)	0.531 (2)	0.923 (5.5)	0.913 (4)	**0.959 (10)**	0.923 (5.5)
Cred	0.625 (1)	0.941 (6)	0.841 (4)	0.983 (8)	**0.990 (10)**	0.937 (5)	0.837 (3)	0.974 (7)	0.724 (2)	0.989 (9)
Diab	0.965 (6)	0.954 (4.5)	0.935 (3)	0.976 (8)	0.979 (9)	0.969 (7)	0.902 (2)	0.954 (4.5)	0.754 (1)	**0.988 (10)**
Germ	0.919 (3)	0.955 (6)	0.955 (6)	0.975 (9)	0.966 (8)	0.891 (2)	0.955 (6)	0.927 (4)	0.852 (1)	**0.981 (10)**
Heart	0.962 (2)	0.974 (5.5)	0.970 (4)	0.984 (8)	**0.996 (10)**	0.974 (5.5)	0.969 (3)	0.975 (7)	0.880 (1)	0.995 (9)
Hepa	0.962 (2)	0.975 (3)	0.990 (6)	0.988 (5)	0.996 (8)	**0.999 (9.5)**	0.991 (7)	0.987 (4)	0.932 (1)	**0.999 (9.5)**
Horse	0.869 (2)	0.908 (7)	0.878 (3)	0.954 (8)	0.981 (9)	0.893 (6)	0.885 (5)	0.880 (4)	0.685 (1)	**0.986 (10)**
Iris	**1.000 (9.5)**	0.994 (7)	0.983 (5)	0.971 (3)	0.977 (4)	0.986 (6)	0.996 (8)	0.866 (2)	0.534 (1)	**1.000 (9.5)**
Monks	0.680 (3)	0.918 (5)	0.986 (9)	0.984 (7)	0.539 (2)	0.985 (8)	0.932 (6)	0.762 (4)	0.423 (1)	**0.996 (10)**
Mush	0.880 (8)	0.874 (7)	0.596 (1)	0.824 (5)	0.759 (4)	0.951 (9)	0.748 (3)	0.847 (6)	0.622 (2)	**0.961 (10)**
Pima	0.904 (3)	0.942 (5)	0.928 (4)	0.957 (9)	0.947 (7.5)	0.945 (6)	0.888 (2)	0.947 (7.5)	0.685 (1)	**0.974 (10)**
Tic	0.733 (2)	0.963 (8)	0.908 (6)	0.920 (7)	0.475 (1)	0.771 (3)	**0.983 (10)**	0.900 (5)	0.843 (4)	0.964 (9)
Wbc	**0.997 (8.5)**	**0.997 (8.5)**	0.976 (3)	0.996 (6)	0.995 (5)	**0.997 (8.5)**	0.756 (1)	0.985 (4)	0.870 (2)	**0.997 (8.5)**
Wdbc	**0.997 (9.5)**	0.995 (6.5)	0.968 (4)	0.987 (5)	0.959 (3)	0.996 (8)	0.834 (2)	0.995 (6.5)	0.489 (1)	**0.997 (9.5)**
Wine	0.941 (8)	**0.947 (10)**	0.885 (6)	0.824 (3)	0.733 (2)	0.944 (9)	0.913 (7)	0.850 (4)	0.371 (1)	0.882 (5)
Average	0.880 (4.9)	0.935 (6.4)	0.900 (4.8)	0.933 (6.4)	0.813 (5.3)	0.914 (6.5)	0.887 (4.7)	0.899 (4.7)	0.712 (2.3)	**0.971 (9.0)**



(a) Datasets 1-8 (b) Datasets 9-16

Fig. 3: Effect of attribute noise on AUC

fore, although the time complexity and operational efficiency of KFRAD are not the best, it is clear from the experimental results and theoretical analysis that KFRAD achieves higher detection accuracy and shows better performance on most datasets.

C. Effect of attribute noise on AUC

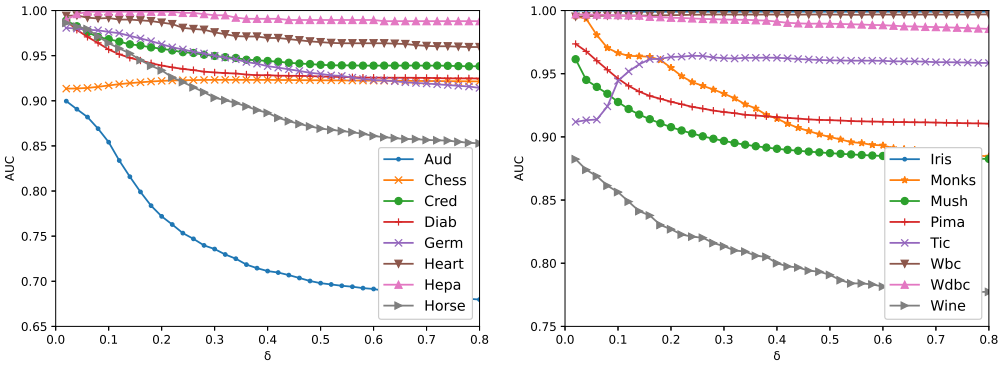
Secondly, we analyze attribute noise's impact on AUC using the method described in [26]. More specifically, error values with a level of $f \cdot 100\%$ are introduced into every attribute. In order to generate the datasets with a noise level of $f \cdot 100\%$, the values of $\lceil f \times n \rceil$ samples at a given attribute are substituted by random values. In this instance, a random value between the maximum and minimum values is chosen to replace the original value for the numerical attribute. When it comes to nominal attributes, the initial value is changed at random with a new nominal attribute value.

We choose to add noise attributes with a different level between 10% to 40% to evaluate the effect of attribute noise on AUC, and the experimental results are shown in Fig. 3. As can be seen from Fig. 3, the AUC values of most datasets fluctuate up and down with the increase of the noise level, but the fluctuation is not large, such as the Aud, Chess, Cred datasets. This shows that KFRAD is robust to attribute noise.

D. Parameter sensitivity analysis

Thirdly, we analyze the sensitivity of KFRAD to δ in order to investigate the effect of the parameter on the model outputs. The figures of AUC relative to δ are shown in Fig. 4. From Fig. 4, it can be seen that on some datasets, the AUC increases with the increase of δ and then gradually levels off, such as Chess and Tic datasets. There are also some datasets where the AUC decreases with increasing δ and then levels off, such as the Heart, Pima dataset. This indicates that once a particular value is reached, the KFRAD parameter δ becomes less sensitive to it. Moreover, on some datasets, the AUC decreases gradually with the increase of δ, such as Aud, Monks, and Wine datasets, which indicates that KFRAD is sensitive to these datasets. Meanwhile, there is no significant change in AUC with increasing δ on datasets such as Iris, Wbc, etc., which suggests that KFRAD is insensitive to these datasets.

From the above analysis, it can be seen that the AUC is greatly affected by the value of the parameter δ on most of



(a) Datasets 1-8 (b) Datasets 9-16

Fig. 4: Variation curve of AUC with δ

the datasets, so determining δ scientifically is a fundamental issue for improving the performance of the KFRAD algorithm.

E. Statistical analysis

Finally, in order to determine the reliability of the results and to prove that the results of the experiment are statistically significant, the statistical analysis of the results is assessed using Friedman's test and Nemenyi's post-hoc test according to the strategy of [25] and [26]. Before applying Friedman's test to any dataset, each algorithm is given a sequence value (1,2,3,...) after being ordered according to AUC from low to high as shown in Table III, where the same sequence value is taken if both algorithms have the same AUC. Friedman's test is then used to examine whether these algorithms differ significantly. Suppose M algorithms are compared on N datasets, the computation of Friedman's test is given in [25], where τ_F satisfies the F distribution with degrees of freedom $(M-1)$ and $(M-1)(N-1)$.

If the original hypothesis that "all algorithms have the same performance" is disproved, it suggests that there are major variances among these algorithms. Nemenyi's post-hoc test must be used in this situation to further distinguish these anomaly detection techniques. The formula for calculating the critical difference (CD) of the average ordinal number was given in [25].

Nemenyi's test figure provides a visual representation of whether the two algorithms are significantly different. In Nemenyi's test figure, the average ordinal number of each algorithm is represented by a point, and the size of the CD is illustrated by a horizontal line segment centered on the point. A collection of algorithms are not significantly different if they are connected by a horizontal line segment. In this study, $M = 10, N = 16$, and τ_F is the F distribution satisfying degrees of freedom 9 and 135. According to Friedman's test, when $\alpha = 0.05$, $\tau_F = 7.4577$ is greater than the critical value of 1.9499. Therefore, the original hypothesis is disproved, which indicates that there is a significant difference between all algorithms, which needs to be further differentiated by Nemenyi's post-hoc test.

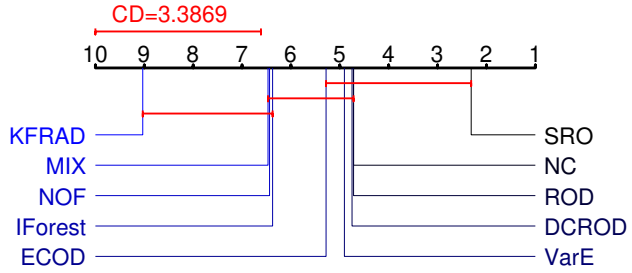


Fig. 5: Nemenyi's test figures on AUC

Fig. 5 is Nemenyi's test figure for the mean ordinal critical difference $CD_{0.05} = 3.3869$ when significant rank $\alpha = 0.05$. Analysis of the figure shows that KFRAD is connected to MIX, NOF, and IForest and not to ECOD, VarE, DCROD, ROD, NC, and SRO. This shows that KFRAD is significantly different from ECOD, VarE, DCROD, ROD, NC, and SRO, but it cannot be proved that KFRAD performs significantly better than MIX, NOF, and IForest.

VII. Conclusions

Aiming at the problem that existing KFRS models are mainly oriented to decision attribute data but cannot effectively handle unlabeled data, this paper proposes a KFRS model for unsupervised anomaly detection. The method makes up for the inadequacy of the existing kernelized fuzzy-rough anomaly detection methods that mainly use intersection operations. In the proposed model, a suitable kernel function is used to compute the kernelized fuzzy relation matrix of attribute subsets and then construct the fuzzy approximation space. Based on this, the proposed KFRS model is applied to the anomaly detection method. For the introduced method, the KFRAD algorithm is designed. The applicability and effectiveness of the algorithm are verified on publicly available datasets.

Currently, the algorithms in this paper are only able to manually set static or fixed parameters. However, different datasets or tasks may have different characteristics and properties, and the proposed method may not be able to adequately cope with dynamically changing data characteristics or task

requirements. By adopting an adaptive approach to determine the optimal parameters, it will be able to better adapt to a variety of complex environments and thereby improve the detection performance of the model. Therefore, in future research, an adaptive approach to determine model parameters efficiently becomes particularly important. In addition, choosing appropriate kernel functions to optimize and extend the model will also be a research direction. In this work, we use the Gaussian kernel to process numerical attribute data, and there are many new kernel functions [32], [43] that can improve the performance of the model or effectively process other attributes. For example, the Match kernel can be used to process nominal attributes, the Histogram intersection kernel is used to process image data, and the Cosine kernel is used in term frequency-inverse document frequency (TF-IDF) attributes of text [32]. We will also investigate certain strategies to combine kernel functions to effectively process multimodality data.

REFERENCES

[1] Z. Li, Y. Zhao, X. Hu, N. Botta, C. Ionescu, and G. Chen, "Ecod: Unsupervised outlier detection using empirical cumulative distribution functions," *IEEE Transactions on Knowledge and Data Engineering*, 2022.

[2] M. Radovanović, A. Nanopoulos, and M. Ivanović, "Reverse nearest neighbors in unsupervised distance-based outlier detection," *IEEE transactions on knowledge and data engineering*, vol. 27, no. 5, pp. 1369–1382, 2014.

[3] B. Tang and H. He, "A local density-based approach for outlier detection," *Neurocomputing*, vol. 241, pp. 171–180, 2017.

[4] M. Carletti, M. Terzi, and G. A. Susto, "Interpretable anomaly detection with diffi: Depth-based feature importance of isolation forest," *Engineering Applications of Artificial Intelligence*, vol. 119, p. 105730, 2023.

[5] J. Li, H. Izakian, W. Pedrycz, and I. Jamal, "Clustering-based anomaly detection in multivariate time series data," *Applied Soft Computing*, vol. 100, p. 106919, 2021.

[6] T. T. Nguyen, "Outlier detection: An approximate reasoning approach," in *Rough Sets and Intelligent Systems Paradigms: International Conference, RSEISP 2007, Warsaw, Poland, June 28-30, 2007. Proceedings 1*, pp. 495–504, Springer, 2007.

[7] Y. Chen, D. Miao, and R. Wang, "Outlier detection based on granular computing," in *Rough Sets and Current Trends in Computing: 6th International Conference, RSCTC 2008 Akron, OH, USA, October 23-25, 2008 Proceedings 6*, pp. 283–292, Springer, 2008.

[8] F. Jiang, Y. Sui, and C. Cao, "Some issues about outlier detection in rough set theory," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4680–4687, 2009.

[9] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *International Journal of General System*, vol. 17, no. 2-3, pp. 191–209, 1990.

[10] Q. Hu, L. Zhang, D. Chen, W. Pedrycz, and D. Yu, "Gaussian kernel based fuzzy rough sets: model, uncertainty measures and applications," *International Journal of Approximate Reasoning*, vol. 51, no. 4, pp. 453–471, 2010.

[11] Q. Hu, D. Yu, W. Pedrycz, and D. Chen, "Kernelized fuzzy rough sets and their applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1649–1667, 2010.

[12] F. Jiang, Y. Sui, and C. Cao, "Outlier detection using rough set theory," in *International Workshop on Rough Sets, Fuzzy Sets, Data Mining, and Granular-Soft Computing*, pp. 79–87, Springer, 2005.

[13] F. Jiang, Y. Sui, and C. Cao, "An information entropy-based approach to outlier detection in rough sets," *Expert Systems with Applications*, vol. 37, no. 9, pp. 6338–6344, 2010.

[14] P. Yang and Q. Zhu, "Finding key attribute subset in dataset for outlier detection," *Knowledge-based systems*, vol. 24, no. 2, pp. 269–274, 2011.

[15] F. Jiang, Y. Sui, and C. Cao, "A hybrid approach to outlier detection based on boundary region," *Pattern recognition letters*, vol. 32, no. 14, pp. 1860–1870, 2011.

[16] A. Albanese, S. K. Pal, and A. Petrosino, "Rough sets, kernel set, and spatiotemporal outlier detection," *IEEE Transactions on knowledge and data engineering*, vol. 26, no. 1, pp. 194–207, 2012.

[17] F. Jiang and Y.-M. Chen, "Outlier detection based on granular computing and rough set theory," *Applied intelligence*, vol. 42, pp. 303–322, 2015.

[18] F. Maciá-Pérez, J. V. Berna-Martinez, A. F. Oliva, and M. A. A. Ortega, "Algorithm for the detection of outliers based on the theory of rough sets," *Decision support systems*, vol. 75, pp. 63–75, 2015.

[19] C. Wang, M. Shao, Q. He, Y. Qian, and Y. Qi, "Feature subset selection based on fuzzy neighborhood rough sets," *Knowledge-Based Systems*, vol. 111, pp. 173–179, 2016.

[20] Y. Chen, D. Miao, and H. Zhang, "Neighborhood outlier detection," *Expert Systems with Applications*, vol. 37, no. 12, pp. 8745–8749, 2010.

[21] Z. Yuan, X. Zhang, and S. Feng, "Hybrid data-driven outlier detection based on neighborhood information entropy and its developmental measures," *Expert Systems with Applications*, vol. 112, pp. 243–257, 2018.

[22] Y. Wang and Y. Li, "Outlier detection based on weighted neighbourhood information network for mixed-valued datasets," *Information Sciences*, vol. 564, pp. 396–415, 2021.

[23] Z. Yuan, H. Chen, T. Li, J. Liu, and S. Wang, "Fuzzy information entropy-based adaptive approach for hybrid feature outlier detection," *Fuzzy Sets and Systems*, vol. 421, pp. 1–28, 2021.

[24] Z. Yuan, H. Chen, T. Li, B. Sang, and S. Wang, "Outlier detection based on fuzzy rough granules in mixed attribute data," *IEEE Transactions on Cybernetics*, vol. 52, no. 8, pp. 8399–8412, 2021.

[25] Z. Yuan, H. Chen, C. Luo, and D. Peng, "Mfgad: Multi-fuzzy granules anomaly detection," *Information Fusion*, vol. 95, pp. 17–25, 2023.

[26] Z. Yuan, B. Chen, J. Liu, H. Chen, D. Peng, and P. Li, "Anomaly detection based on weighted fuzzy-rough density," *Applied Soft Computing*, vol. 134, p. 109995, 2023.

[27] B. Moser, "On the t-transitivity of kernels," *Fuzzy Sets and Systems*, vol. 157, no. 13, pp. 1787–1796, 2006.

[28] D. Chen, Q. He, and X. Wang, "Frsvms: Fuzzy rough set based support vector machines," *Fuzzy Sets and Systems*, vol. 161, no. 4, pp. 596–607, 2010.

[29] Y. Qu, C. Shang, Q. Shen, N. Mac Parthaláin, and W. Wu, "Kernel-based fuzzy-rough nearest neighbour classification," in *2011 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2011)*, pp. 1523–1529, IEEE, 2011.

[30] Y. Li, Y. Lin, J. Liu, W. Weng, Z. Shi, and S. Wu, "Feature selection for multi-label learning based on kernelized fuzzy rough sets," *Neurocomputing*, vol. 318, pp. 271–286, 2018.

[31] D. S. Yeung, D. Chen, E. C. Tsang, J. W. Lee, and W. Xizhao, "On the generalization of fuzzy rough sets," *IEEE Transactions on fuzzy systems*, vol. 13, no. 3, pp. 343–361, 2005.

[32] Q. Hu, L. Zhang, Y. Zhou, and W. Pedrycz, "Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 26, no. 1, pp. 226–238, 2017.

[33] X. Li, J. Lv, and Z. Yi, "Outlier detection using structural scores in a high-dimensional space," *IEEE transactions on cybernetics*, vol. 50, no. 5, pp. 2302–2310, 2018.

[34] J. Huang, Q. Zhu, L. Yang, and J. Feng, "A non-parameter outlier detection algorithm based on natural neighbor," *Knowledge-Based Systems*, vol. 92, pp. 71–77, 2016.

[35] K. Li, X. Gao, S. Fu, X. Diao, P. Ye, B. Xue, J. Yu, and Z. Huang, "Robust outlier detection based on the changing rate of directed density ratio," *Expert Systems with Applications*, vol. 207, p. 117988, 2022.

[36] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *2008 eighth ieee international conference on data mining*, pp. 413–422, IEEE, 2008.

[37] H. Xu, Y. Wang, Y. Wang, and 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)*, pp. 1408–1413, IEEE, 2019.

[38] X. Li, J. Lv, and Z. Yi, "An efficient representation-based method for boundary point and outlier detection," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 1, pp. 51–62, 2016.

[39] Y. Almardeny, N. Boujnah, and F. Cleary, "A novel outlier detection method for multivariate data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 9, pp. 4052–4062, 2020.

[40] C. You, D. P. Robinson, and 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*, pp. 3395–3404, 2017.

[41] C. Liu, Z. Yuan, B. Chen, H. Chen, and D. Peng, "Fuzzy granular anomaly detection using markov random walk," *Information Sciences*, vol. 646, p. 119400, 2023.

[42] C. Liu, D. Peng, H. Chen, and Z. Yuan, "Attribute granules-based object entropy for outlier detection in nominal data," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108198, 2024.
[43] V. H. Moghaddam and J. Hamidzadeh, "New hermite orthogonal polynomial kernel and combined kernels in support vector machine classifier," *Pattern Recognition*, vol. 60, pp. 921–935, 2016.

Yan Wu is currently pursuing the B.S. degree with the College of Computer Science, Sichuan University, Chengdu, China.

His research interests include granular computing, anomaly detection, and outlier detection.

Sihan Wang is currently pursuing the B.S. degree with the College of Computer Science, Sichuan University, Chengdu, China.

Her research interests include granular computing, granular-ball computing, and outlier detection.

Hongmei Chen (Member, IEEE) received the M.Sc. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2000. She received the Ph.D. degree from the Southwest Jiaotong University, Chengdu, China, in 2013. She is currently a Professor at the School of Computing and Artificial Intelligence, Southwest Jiaotong University.

Her research interests include the areas of data mining, pattern recognition, fuzzy sets, and rough sets.

Dezhong Peng (Member, IEEE) received the B.Sc. degree in applied mathematics, and the M.Sc. and Ph.D. degrees in computer software and theory from the University of Electronic Science and Technology of China, Chengdu, China, in 1998, 2001, and 2006, respectively. From 2001 to 2007, he was with the University of Electronic Science and Technology of China as an Assistant Lecturer and a Lecturer. He was a Postdoctoral Research Fellow with the School of Engineering, Deakin University, Geelong, VIC, Australia, from 2007 to 2009. He is currently a Professor at the College of Computer Science, Sichuan University, Chengdu, China.

His research interests include neural networks and signal processing.

Zhong Yuan received the M.Sc. degree in mathematics from Sichuan Normal University, Chengdu, China, in 2018. He received the Ph.D. degree from the Southwest Jiaotong University, Chengdu, China, in 2022. He is currently a Distinguished Associate Researcher at the College of Computer Science, Sichuan University.

His research interests include granular computing, uncertainty information processing, and outlier detection.