Information Fusion 95 (2023) 17–25

Contents lists available at ScienceDirect

Information Fusion

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

Full length article

MFGAD: Multi-fuzzy granules anomaly detection

Zhong Yuan [a], Hongmei Chen [b], Chuan Luo [a], Dezhong Peng [a,c,d,*]

[a] College of Computer Science, Sichuan University, Chengdu 610065, China
[b] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu 611756, China
[c] Sichuan Zhiqian Technology Co., Ltd., Chengdu 610041, China
[d] Chengdu Ruibei Yingte Information Technology Ltd. Company, Chengdu 610094, China

ARTICLE INFO

Dataset link: https://github.com/BElloney/Outlier-detection

Keywords:
Granular computing
Fuzzy rough set theory
Unsupervised anomaly detection
Multi-granularity
Hybrid data

ABSTRACT

Unsupervised anomaly detection is an important research direction in the process of unsupervised knowledge acquisition. It has been successfully applied in many fields, such as online fraud identification, loan approval, and medical diagnosis. Multi-granularity thinking is an effective information fusion method for solving problems in a multi-granular environment, which allows people to understand and analyze problems from multiple perspectives. However, there are few studies on building anomaly detection models using the idea of multi-fuzzy granules. To this end, this paper constructs a multi-fuzzy granules anomaly detection method by using a fuzzy rough computing model. In this method, a hybrid metric is first used to calculate the fuzzy relations. Then, two ranking sequences are constructed based on the significance of attributes. Furthermore, forward and reverse multi-fuzzy granules are constructed to define anomaly scores based on the ranking sequences. Finally, a multi-fuzzy granules-based anomaly detection algorithm is designed to detect anomalies. The experimental results compared with existing algorithms show the effectiveness of the proposed algorithm.

1. Introduction

In recent years, the discovery of anomalies (or outliers) from data has attracted increasing attention because of its wide range of potential applications. Its purpose is to find observations that are different from expected [1]. In most unsupervised data mining research, outliers are usually discarded as noise [2,3]. For example, Jiang et al. used outlier detection techniques to ensure that the selected initial cluster centers are not outliers [2]. The idea is that outliers should not be selected as initial cluster centers and need to be removed. However, finding outliers belonging to a minority class may be more interesting than finding normal points in many applied studies, such as fraud detection [4], process monitoring [5], and sensor networks [6]. Therefore, effectively detecting such abnormal samples has important research significance and application value in real life.

Granular Computing (GrC), an important tool for knowledge discovery, is a novel theory and method for simulating the thinking patterns of humans in solving large-scale complex problems [7]. Rough computing theory is an outstanding computational model of GrC theory, which to some extent emphasizes the importance of the concept of granularity. Therefore, it has become a popular mathematical framework in GrC theory and has been used in a variety of applications. However, in the rough computing method, the learning model is established based on the equivalence relation, so its learning model is only suitable for nominal attributes. Before building a learning model, the numerical attribute data must be discretized, which may change the intrinsic structure of the data itself, thereby affecting the accuracy of detection. To effectively handle numerical or hybrid (or mixed) data, fuzzy rough computing theory has been widely discussed and researched, and has been successfully applied to feature reduction [8–10], dimensionality reduction [11,12], multiple attribute group decision-making [13], stock forecasting [14], and three-way classification [15]. Nevertheless, there are few studies on fuzzy rough computing models for unsupervised outlier detection [16–18], and further research is still needed.

Information fusion is the process of combining multiple pieces of information from different sources in order to create a more complete and accurate understanding of a situation or event, such as multi-source [19], multi-modal [20], multi-granularity [21], etc. For example, Qian et al. extended the classical rough computing model to a multi-granulation rough computing model [21], which utilizes multiple equivalence relations in a universe to define upper and lower approximation sets. According to the idea of the multi-granularity rough set model, a series of related models are proposed, such as adaptive multi-granulation decision theory rough set [22], covering-based multi-granulation fuzzy rough set [23] and multi-granulation supertrust model [24]. The idea of multi-granularity enables people to understand and analyze problems from multiple perspectives, and

* Corresponding author.
E-mail addresses: yuanzhong@scu.edu.cn (Z. Yuan), hmchen@swjtu.edu.cn (H. Chen), cluo@scu.edu.cn (C. Luo), pengdz@scu.edu.cn (D. Peng).

https://doi.org/10.1016/j.inffus.2023.02.007
Received 11 October 2022; Received in revised form 3 February 2023; Accepted 5 February 2023
Available online 8 February 2023
1566-2535/© 2023 Elsevier B.V. All rights reserved.

provides an effective information fusion method for solving problems in a multi-granular environment [25,26]. Specifically, the idea of multi-granularity information fusion induces different information granular structures through different subsets of attributes. Then, their effective information is fully fused to construct subsequent learning models. However, to the best of our knowledge, there are few researchers using the idea of multi-fuzzy granules to construct anomaly detection models.

Based on the above discussion, this paper constructs a multi-fuzzy granules-based anomaly detection model. The model takes advantage of the fuzzy rough computing model's ability to deal with fuzzy and uncertain data effectively. First, in order to efficiently handle hybrid data, a hybrid metric is adopted to compute the fuzzy relations. Second, the significance of attributes is defined based on fuzzy entropy to construct attribute sequences and attribute subset sequences. Furthermore, based on the two sequences, i.e., forward and reverse multi-fuzzy granules, a novel concept of multi-fuzzy granules anomaly is proposed to characterize the degree of anomaly of the samples. Finally, the corresponding Multi-Fuzzy Granules Anomaly Detection (MFGAD) algorithm is designed. Comparative experimental results show the effectiveness and applicability of the proposed algorithm.

The rest of the paper is organized as follows. The second section briefly describes the work related to this study. Section 3 reviews some preliminary knowledge about fuzzy rough computing models. Section 4 proposes a multi-fuzzy granules anomaly detection model, including methods and algorithms. Section 5 is the experiments. Finally, Section 6 summarizes the work of this paper.

2. Related work

In this section, some rough computing methods for anomaly detection are reviewed, including the classical rough computing methods and extended rough computing methods.

2.1. Classical rough computing method

For the purpose of effectively detecting anomalies in nominal attribute data, scholars have proposed some anomaly detection methods based on rough computing models. For example, Jiang et al. earlier developed a new method for the definition and detection of outliers based on rough computing theory [27]. The main idea is that a sample in the boundary region is more likely to be an outlier than a sample in the lower approximation. Further, they studied the rough membership function-based outliers [28], the sequence-based outliers [29], the information entropy-based outliers [30], and the boundary and distance-based outliers [31]. Chen et al. proposed a novel concept of outliers based on granular computing [32]. Aiming at the problem of anomaly detection in unlabeled spatiotemporal data, Albanese et al. put forward a rough computing models-based anomaly detection approach [33]. Jiang and Chen proposed a new outlier detection method from the perspective of GrC and rough set theory [34]. Maciá-Pérez et al. built a basic theory of rough computing for anomaly detection and proposed an algorithm that can detect anomalies in a large amount of information [35]. In [36], an anomaly detection algorithm was developed by using approximation accuracy entropy under the rough computing framework. Singh et al. presented an anomaly detection method for large-scale data streams using rough computing models [37]. However, the above rough computing methods are only applicable to nominal attribute data because their detection model is based on equivalence relations.

2.2. Extended rough computing method

In response to the shortcomings of the above rough computing methods, scholars have proposed some extended rough computing methods. For example, Chen et al. employed the neighborhood rough computing model as a unified framework to understand and implement anomaly detection [38]. Yuan et al. studied the anomaly detection method of neighborhood entropy based on hybrid data-driven [39]. This method extends the traditional detection methods based on distance and rough sets, and is suitable for three data types. Aiming at the problem of anomaly detection in mixed data, a detection method based on multigranulation relative entropy is proposed in neighborhood rough set theory [40]. Wang and Li proposed an outlier detection method for mixed-value data based on a weighted network model [41], which constructs a weighted neighborhood information network to represent a mixed-value attribute dataset by considering neighborhood relations and similarities between samples. Using a fuzzy approximation space with fuzzy similarity relations, Yuan et al. constructed a hybrid feature outlier detection method based on fuzzy information entropy [17]. In [16], fuzzy rough computing models are introduced to deal with anomaly detection in mixed data. However, to our knowledge, the research on using multi-fuzzy granule information to construct anomaly detection has not been reported.

3. Preliminaries

In fuzzy rough computing, a data table without decision can be imported into an information system, which is represented by a two-tuple $IS = \langle S, A \rangle$. Herein, S is a non-empty finite sample set; A is a non-empty finite conditional attribute set; for any $s \in S$ and $a \in A$, $f_a(s)$ denotes the value of s with respect to attribute a.

Definition 1. Let $S = \{s_1, s_2, \ldots, s_n\}$, if $\widetilde{X}$ is a mapping from S to $[0,1]$, i.e.,

$$\widetilde{X} : S \to [0,1], \tag{1}$$

then $\widetilde{X}$ is called a fuzzy set on S.

Let $\widetilde{X}$ and $\widetilde{Y}$ be two fuzzy sets on S. For any $s \in S$, some related operations are as follows.

(1) Inclusion: $\widetilde{X}(s) \le \widetilde{Y}(s) \Rightarrow \widetilde{X} \subseteq \widetilde{Y}$;
(2) Intersection: $(\widetilde{X} \cap \widetilde{Y})(s) = \widetilde{X}(s) \wedge \widetilde{Y}(s)$;
(3) Union: $(\widetilde{X} \cup \widetilde{Y})(s) = \widetilde{X}(s) \vee \widetilde{Y}(s)$.

Definition 2. A fuzzy relation $\widetilde{R}$ about S refers to a fuzzy set on $S \times S$, i.e.,

$$\widetilde{R} : S \times S \to [0,1]. \tag{2}$$

Let $\widetilde{R}$ be a fuzzy relation on S. For any $s, t \in S$, if $\widetilde{R}$ satisfies (1) reflexivity ($\widetilde{R}(s,s) = 1$) and (2) symmetry ($\widetilde{R}(s,t) = \widetilde{R}(t,s)$), then $\widetilde{R}$ is a fuzzy similarity relation on S. The fuzzy granule family $G(\widetilde{R})$ of S induced by $\widetilde{R}$ is defined as $G(\widetilde{R}) = \{[s_1]_{\widetilde{R}}, [s_2]_{\widetilde{R}}, \ldots, [s_n]_{\widetilde{R}}\}$, where $[s_i]_{\widetilde{R}} = \frac{\widetilde{R}(s_i,s_1)}{s_1} + \frac{\widetilde{R}(s_i,s_2)}{s_2} + \cdots + \frac{\widetilde{R}(s_i,s_n)}{s_n} = (\widetilde{R}(s_i,s_1), \widetilde{R}(s_i,s_2), \ldots, \widetilde{R}(s_i,s_n))$ is a fuzzy granule induced by $\widetilde{R}$.

To better handle the data, Dubois and Prade proposed the following concept of fuzzy rough sets [42].

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

$$\underline{\widetilde{R}}\widetilde{Y}(s) = \inf_{t \in S} \{(1 - \widetilde{R}(s,t)) \vee \widetilde{Y}(t)\}; \tag{3}$$

$$\overline{\widetilde{R}}\widetilde{Y}(s) = \sup_{t \in S} \{\widetilde{R}(s,t) \wedge \widetilde{Y}(t)\}. \tag{4}$$

Table 1
An information system.

S	a_1	a_2	a_3
s_1	c	4	0.7
s_2	a	7	0.4
s_3	c	1	0.6
s_4	a	2	0.3
s_5	a	8	0.5
s_6	b	10	0.8

Fuzzy rough set model combines the advantages of fuzzy sets and rough sets, which can effectively deal with fuzzy or imprecise data. Fuzzy granules, as the basic granule unit of fuzzy rough set theory, include the fuzzy information of each sample. Next, we will construct a multi-fuzzy granules-based anomaly detection model by analyzing the variation difference between fuzzy granules regarding the sequence of attribute subsets.

4. Multi-fuzzy granules anomaly detection

This section proposes a multi-fuzzy granules anomaly detection, including a detection model and its corresponding algorithm.

4.1. Detection model

In the data processing of information systems, the data usually have the difference in the order of magnitude and dimension. In order to obtain accurate data processing results, the min–max standardization is utilized to make the range $[0,1]$ for all numerical attributes.

In order to efficiently handle numerical and mixed attribute data, a hybrid fuzzy similarity measure is utilized to calculate the fuzzy similarity relation [17], i.e., the fuzzy similarity degree $\widetilde{R}_a(s,t)$ between samples s and t with respect to $a \in A$ is calculated as

$$\widetilde{R}_a(s,t)=\begin{cases}1, f_a(s)=f_a(t) \text{ and } a \text{ is nominal;}\\ 0, f_a(s)\neq f_a(t) \text{ and } a \text{ is nominal;}\\ 1-|f_a(s)-f_a(t)|, |f_a(s)-f_a(t)|\le \varepsilon_a \text{ and } a \text{ is numerical;}\\ 0, |f_a(s)-f_a(t)|>\varepsilon_a \text{ and } a \text{ is numerical,}\end{cases} \tag{5}$$

where ε_a is the adjustable fuzzy radius, which is calculated by $\varepsilon_a = \frac{std(a)}{\delta}$, where $std(a)$ is the standard deviation of the attribute value on a and δ is an adjustable parameter.

Let $\widetilde{R}_E$ be a fuzzy relation on $E \subseteq A$, which can be represented by the following fuzzy relation matrix, i.e., $M(\widetilde{R}_E) = (r^E_{ij})_{n\times n}$ where $r^E_{ij} = \widetilde{R}_E(s_i,s_j)$.

Example 1. Given an $IS = \langle S, A\rangle$, as shown in Table 1, where $S = \{s_1,s_2,s_3,s_4,s_5,s_6\}$ and $A = \{a_1,a_2,a_3\}$, the 2nd column is nominal data, and the 3rd and 4th columns are both numeric data. First, the raw numerical attribute data is standardized to [0,1] by using the min–max standardization method. Let $\delta = 1$, then the corresponding fuzzy radii of attributes a_2 and a_3 are $\varepsilon_{a_2} \approx 0.3610$ and $\varepsilon_{a_3} \approx 0.3416$, respectively.

According to Formulas (5), the fuzzy relation matrices of each attribute in A are computed as

$$M(\widetilde{R}_{a_1})=\begin{bmatrix}1&0&1&0&0&0\\0&1&0&1&1&0\\1&0&1&0&0&0\\0&1&0&1&1&0\\0&1&0&1&1&0\\0&0&0&0&0&1\end{bmatrix};$$

$$M(\widetilde{R}_{a_2})=\begin{bmatrix}1&0.667&0.667&0.778&0&0\\0.667&1&0&0&0.889&0.667\\0.667&0&1&0.889&0&0\\0.778&0&0.889&1&0&0\\0&0.889&0&0&1&0.778\\0&0.667&0&0&0.778&1\end{bmatrix};$$

$$M(\widetilde{R}_{a_3})=\begin{bmatrix}1&0&0.8&0&0&0.8\\0&1&0&0.8&0.8&0\\0.8&0&1&0&0.8&0\\0&0.8&0&1&0&0\\0&0.8&0.8&0&1&0\\0.8&0&0&0&0&1\end{bmatrix}.$$

Definition 4. The fuzzy granule family $G(\widetilde{R}_E)$ of S induced by $\widetilde{R}_E$ is defined as

$$G(\widetilde{R}_E) = \{[s_1]_{\widetilde{R}_E}, [s_2]_{\widetilde{R}_E}, \dots, [s_n]_{\widetilde{R}_E}\}, \tag{6}$$

where $[s_i]_{\widetilde{R}_E} = \frac{r^E_{i1}}{s_1} + \frac{r^E_{i2}}{s_2} + \cdots + \frac{r^E_{in}}{s_n} = (r^E_{i1}, r^E_{i2}, \dots, r^E_{in})$ is a fuzzy granule induced by $\widetilde{R}_E$. Without causing confusion, this article also uses E instead of $\widetilde{R}_E$.

Let $E = \{a_{k_1}, a_{k_2}, \dots, a_{k_h}\}(1 \le h \le m) \subseteq A = \{a_1,a_2,\dots,a_m\}$. Obviously, $[s_i]_E$ is a fuzzy set with respect to $\widetilde{R}_E$ on S. For the determination of $\widetilde{R}_E(s_i,s_j)$, there are various calculation methods [43]. In this paper, the conjunction method is adopted, i.e., $\widetilde{R}_E(s_i,s_j) = \bigwedge_{l=1}^{h} \widetilde{R}_{a_{k_l}}(s_i,s_j)$. The cardinality of $[s_i]_E$ is defined as $|[s_i]_E| = \sum_{j=1}^n r^E_{ij}$. Obviously, $1 \le \left|[s_i]_E\right| \le n$.

Property 1. *It is easy to prove that the following conclusions hold.*

(1) $E \subseteq A \Rightarrow \widetilde{R}_A \subseteq \widetilde{R}_E$;
(2) $E \subseteq A \Rightarrow [s_i]_A \subseteq [s_i]_E$.

Through the above normalization, hybrid metric and adjustable fuzzy radius, a corresponding fuzzy information system is obtained. Next, anomaly scores are constructed for detecting anomalies.

In order to solve the problem of quantitative measurement of fuzzy information, Hu et al. introduced information entropy into fuzzy rough set theory, and proposed the following new concept of fuzzy entropy [44].

Definition 5. The fuzzy entropy with respect to $\widetilde{R}_E$ is defined by

$$FE(E) = FE(\widetilde{R}_E) = -\frac{1}{n}\sum_{i=1}^{n}\log_2\frac{|[s_i]_E|}{|S|}. \tag{7}$$

The fuzzy entropy defined above can reflect the uncertainty of information systems.

Property 2. *If $\widetilde{R}_{a_s} \subseteq \widetilde{R}_{a_t}$, then $FE(\widetilde{R}_{a_s}) \ge FE(\widetilde{R}_{a_t})$.*

Proof. Let $\widetilde{R}_{a_s} \subseteq \widetilde{R}_{a_t}$. So for $\forall r^{a_s}_{ij}, r^{a_t}_{ij}$, there is $\forall r^{a_s}_{ij} \le r^{a_t}_{ij}$. So there is $|[s_i]_{a_s}| = \sum_{j=1}^n r^{a_s}_{ij} \le \sum_{j=1}^n r^{a_t}_{ij} = |[s_i]_{a_t}|$, i.e., $-\frac{1}{n}\sum_{i=1}^n \log_2\frac{|[s_i]_{a_s}|}{|S|} \ge -\frac{1}{n}\sum_{i=1}^n \log_2\frac{|[s_i]_{a_t}|}{|S|}$. Therefore, there is $FE(\widetilde{R}_{a_s}) \ge FE(\widetilde{R}_{a_t})$.

Property 2 shows that fuzzy entropy varies monotonically with the fuzzy relation matrix of a single attribute. The larger the fuzzy relation matrix of an attribute, the smaller the distinguishing ability and the smaller fuzzy entropy. Conversely, the larger the distinguishing ability, the larger fuzzy entropy. The greater the fuzzy entropy of an attribute, the greater its distinguishing ability and the more important it is. Therefore, it can describe the significance of an attribute.

Definition 6. The significance of a_k is defined as

$$sig(a_k) = FE(\{a_k\}). \tag{8}$$

Example 2. Continued Example 1. According to Definition 5, the significance of each attribute is computed as
$Sig(a_1) = FE(\{a_1\}) = -\frac{1}{|S|}\sum_{i=1}^{n}\log_2\frac{|[s_i]_{a_1}|}{|S|} = -(\frac{1}{6}\log_2\frac{2}{6} + \frac{1}{6}\log_2\frac{3}{6} + \frac{1}{6}\log_2\frac{2}{6} + \frac{1}{6}\log_2\frac{3}{6} + \frac{1}{6}\log_2\frac{3}{6} + \frac{1}{6}\log_2\frac{1}{6}) \approx 1.4591$. Similarly, there are $Sig(a_2) = FE(\{a_2\}) \approx 1.1185$, $Sig(a_3) = FE(\{a_3\}) \approx 1.3833$.

If the significance of each attribute in A has been calculated, then all the attributes in A can be sorted by significance to get an attribute sequence.

Definition 7. The attribute sequence is constructed as

$$AS = \langle a'_1, a'_2, \dots, a'_m\rangle, \tag{9}$$

where $Sig(a'_k) \le Sig(a'_{k+1})$.

Further, if we start with the attribute set A, the attribute with the least (or greatest) significance in A is gradually removed, until finally a set containing only one attribute is obtained. In this way, two kinds of attribute subset sequences can be obtained.

Definition 8. The forward and reverse sequences of attribute subsets are respectively constructed as

$$FS = \langle A_1, A_2, \dots, A_m\rangle; \tag{10}$$

$$RS = \langle A'_1, A'_2, \dots, A'_m\rangle, \tag{11}$$

where $A_k \subseteq A, A_1 = A, A_m = \{a'_m\}$ and $A_{k+1} = A_k - \{a'_k\}$; $A'_k \subseteq A, A'_1 = A, A'_m = \{a'_1\}$ and $A'_{k+1} = A'_k - \{a'_{m-k+1}\}$.

Example 3. Continued Example 2. By Definition 7, the attribute sequence is constructed as $S = \langle a'_1, a'_2, a'_3\rangle = \langle a_2, a_3, a_1\rangle$. Further, by Definition 8, the following two attribute subset sequences can be constructed as $FS = \langle A_1, A_2, A_3\rangle = \langle\{a_1, a_2, a_3\}, \{a_1, a_3\}, \{a_1\}\rangle$ and $RS = \langle A'_1, A'_2, A'_3\rangle = \langle\{a_1, a_2, a_3\}, \{a_2, a_3\}, \{a_2\}\rangle$.

From Definition 8, each set of two attribute sequences can determine a fuzzy relation, which results in a multi-fuzzy granules structure about all fuzzy relations.

Definition 9. The forward and reverse multi-fuzzy granules of s are respectively constructed as

$$FMG = \langle [s]_{A_1}, [s]_{A_2}, \dots, [s]_{A_m}\rangle; \tag{12}$$

$$RMG = \langle [s]_{A'_1}, [s]_{A'_2}, \dots, [s]_{A'_m}\rangle. \tag{13}$$

Property 3. *Obviously, the following properties hold.*

$$[s]_{A_1} \subseteq [s]_{A_2} \subseteq \cdots \subseteq [s]_{A_m}; \tag{14}$$

$$[s]_{A'_1} \subseteq [s]_{A'_2} \subseteq \cdots \subseteq [s]_{A'_m}. \tag{15}$$

It is often necessary to construct an anomaly score to characterize the degree of anomaly of a sample in anomaly detection. However, methods for considering the idea of multi-fuzzy granules are rare. Herein, multi-fuzzy granules information is integrated to construct anomaly scores.

Definition 10. The multi-fuzzy granules-based anomaly score of s is defined as

$$Score(s) = 1 - W(s)\sqrt{\frac{\sum_{k=2}^{m}\left(\frac{|[s]_{A_k}|-|[s]_{A_{k-1}}|}{|[s]_{A_k}|} + \frac{|[s]_{A'_k}|-|[s]_{A'_{k-1}}|}{|[s]_{A'_k}|}\right)}{2(m-1)}}, \tag{16}$$

where the weight $W(s) = \frac{\sum_{k=1}^{m}\sqrt{\frac{|[s]_{a_k}|}{|S|}}}{|A|}$.

In the above definition, the anomaly score is defined by fusing the mean value of the ratio of the relative differences of multi-fuzzy granules and the corresponding weights. Here, the smaller the mean value of the ratio of the relative difference of multi-fuzzy granules is, the more anomalous the sample s may be. In this case, we should assign a relatively larger anomaly score to the sample s. In addition, $W(s)$ reflects the opinion that anomalies tend to belong to a minority class of the sample, and minority samples are more likely to be anomalous than majority samples. In particular, $Score(s)$ is inversely proportional to its $W(s)$, i.e., the smaller $W(s)$, the more likely it is that s is an abnormal sample.

Definition 11. Given a threshold ν. For any $s \in S$, if $Score(s) > \nu$, then s is called a multi-fuzzy granules-based anomaly in S.

Example 4. Continued Example 3. For $s_1 \in S$, we have

$$\begin{aligned} FMG &= \langle [s_1]_{A_1}, [s_1]_{A_2}, [s_1]_{A_3}\rangle \\ &= \langle (1, 0, 0.667, 0, 0, 0), (1, 0, 0.8, 0, 0, 0), (1, 0, 1, 0, 0, 0)\rangle; \end{aligned}$$

$$\begin{aligned} RMG &= \langle [s_1]_{A'_1}, [s_1]_{A'_2}, [s_1]_{A'_3}\rangle \\ &= \langle (1, 0, 0.667, 0, 0, 0), (1, 0, 0.667, 0, 0, 0), (1, 0.667, 0.667, 0.778, 0, 0)\rangle. \end{aligned}$$

Further, the weights can be calculated as $W(s_1) = \frac{\sqrt{\frac{2}{6}}+\sqrt{\frac{3.112}{6}}+\sqrt{\frac{2.6}{6}}}{3} \approx 0.428$.

Algorithm 1: MFGAD

Input: $IS = \langle S, A\rangle$, δ
Output: $Score$

```
Score ← ∅;
for k ← 1 to |A| do
    Compute M(R̃_{a_k}) by Equation (5);
    Compute Sig(a_k) by Definition 6;
end
Determine AS by Definition 7;
Construct FS and RS by Definition 8;
for k ← 1 to |A| do
    Compute M(R̃_{A_k}) and M(R̃_{A'_k});
end
for i ← 1 to |S| do
    Compute FMG and RMG by Definition 9;
    Compute Score(s_i) by Definition 10;
end
return Score.
```

Therefore, according to Definition 10, the multi-fuzzy granules-based anomaly score of s_1 is calculated as follows.

$$\begin{aligned} Score(s_1) &= 1 - W(s_1)\sqrt{\frac{\sum_{k=2}^{m}\left(\frac{|[s_1]_{A_k}|-|[s_1]_{A_{k-1}}|}{|[s_1]_{A_k}|} + \frac{|[s_1]_{A'_k}|-|[s_1]_{A'_{k-1}}|}{|[s_1]_{A'_k}|}\right)}{2(m-1)}} \\ &= 1 - 0.428 \times \sqrt{\frac{(\frac{1.8-1.667}{1.8} + \frac{1.667-1.667}{1.667}) + (\frac{2-1.8}{2} + \frac{3.112-1.667}{3.112})}{2(3-1)}} \\ &\approx 0.829. \end{aligned}$$

Similarly, we can get $Score(s_2) \approx 0.770$, $Score(s_3) \approx 0.856$, $Score(s_4) \approx 0.749$, $Score(s_5) \approx 0.804$, $Score(s_6) \approx 0.888$.

By Definition 11, let $\nu = 0.850$, we have $Score(s_1)$, $Score(s_2)$, $Score(s_4)$, $Score(s_5) < \nu$, $Score(s_3)$, $Score(s_6) > \nu$. It can be seen that the samples s_3 and s_6 are both multi-fuzzy granules-based anomalies.

Table 2
Basic information of the experimental datasets.

ID	Datasets	Abbr.	Attributes	Samples	Anomalies	Type
1	Annthyroid	Ann	6	7200	534	Numeric
2	Cardio	Card	21	1831	176	Numeric
3	CreditA_plus_42_variant1	Cred	15	425	42	Hybrid
4	Diabetes_tested_positive_26_variant1	Diab	8	526	26	Numeric
5	German_1_14_variant1	Germ	20	714	14	Hybrid
6	Heart_2_16_variant1	Heart	13	166	16	Hybrid
7	Hepatitis_2_9_variant1	Hepa	19	94	9	Hybrid
8	Ionosphere_b_24_variant1	Iono	34	249	24	Numeric
9	Mushroom_p_221_variant1	Mush	22	4429	221	Nominal
10	Pima_TRUE_55_variant1	Pima	9	555	55	Numeric
11	Sonar_M_10_variant1	Sonar	60	107	10	Numeric
12	Thyroid	Thyr	6	3772	93	Numeric
13	Vertebral	Vert	6	240	30	Numeric
14	Wbc_malignant_39_variant1	Wbc	9	483	39	Numeric
15	Wdbc_M_39_variant1	Wdbc	31	396	39	Numeric
16	Wine	Wine	13	129	10	Numeric

4.2. Detection algorithm

In the previous subsection, a multi-fuzzy granules-based anomaly detection model is presented, including the construction of information system, the construction of detection model, and the determination of anomalies. This subsection mainly designs the corresponding MFGAD algorithm.

In Algorithm 1, we first initialize $Score$ to $\emptyset$. In Steps 2–5, the fuzzy relation matrix and significance of each attribute are calculated by a "for" loop. In Steps 6–7, AS, FS and RS are constructed sequentially. In Steps 8–10, the fuzzy relation matrix for each attribute subset is calculated by a "for" loop. In Steps 11–14, multi-fuzzy granules are computed by a "for" loop to construct anomaly scores for each sample. Through analysis, the total number of rounds of Algorithm 1 is $|A| \times |S| \times |S| + |A| + |S|$. Therefore, the time complexity of Algorithm 1 is $O(|A||S|^2)$.

5. Experiments

In this section, we will conduct experiments to analyze the effectiveness and parameter sensitivity of the proposed method. To test its effectiveness, we will compare algorithm MFGAD with some existing algorithms on a public dataset. These comparison algorithms are DIStance (DIS)-based algorithm [45], Connectivity-based Outlier Factor (COF) [46], INFLuenced Outlierness (INFLO) [47], Local Distance-based Outlier Factor (LDOF) [48], Local Outlier Probabilities (LoOP) [49], Outlier Detection using Indegree Number (ODIN) [50], Directed density ratio Changing Rate-based Outlier Detection (DCROD) [51], Empirical Cumulative-based Outlier Detection (ECOD) [52], Rotation-based Outlier Detection (ROD) [53], Granular Computing (GrC)-based outliers [32], SEQuence (SEQ)-based method [29], Information Entropy (IE)-based outliers [30], Weighted Neighborhood Information Network-based Outlier Detection (WNINOD) [41]. For the parameter sensitivity analyses, we will give the curve of the performance of the proposed algorithm with δ.

5.1. Experimental setups

We download sixteen datasets for anomaly detection from relevant public web pages[1,2] for experiments. The relevant experimental datasets are described in Table 2.

In the experiment, for algorithms COF, INFLO, LDOF, LoOP, ODIN, and DCROD, the optimal value of their parameters are calculated in the range of $[1, 60]$ and the step size of 1. Furthermore, algorithms GrC, SEQ, and IE are generally only applicable to nominal attribute data. Thus, the FCM (Fuzzy C-Means) discretization method is used to convert the numeric attribute values into nominal attributes, where the number of intervals is 3. For algorithms DIS, COF, INFLO, LDOF, LoOP, ODIN, and DCROD, the Euclidean distance metric is used and all different nominal attributes in the dataset are replaced with different integer values. Further, all attribute values are normalized to $[0, 1]$ interval using the min–max normalization method. The parameter adjustment range of WNINOD is $[1, 10]$ and the adjustment step size is set to 1. Since the output of WNINOD is inlier scores, it will be inconvenient for experimental comparison. For ease of comparison, the inverse of all inlier scores of its output is taken as the anomaly score. For the algorithm proposed in this paper, the optimal value of δ is computed in $[0.1, 2]$ with a step size of 0.1. Furthermore, in traditional distance-based detection methods, an anomaly score is defined using the strategy in [29] to indicate the degree of anomaly of each sample.

To comprehensively evaluate the performance of the algorithm, the Receiver Operating Curve (ROC) and the Area Under Curve (AUC) indexes are adopted in the experiments [54]. In general, most outlier detection algorithms eventually output an anomaly score for each sample and a score threshold is set to determine the anomalies. Given a score threshold v, let OS_d denote the currently detected outlier set, which is determined by

$$OS_d = \{s | Score(s) > v\}. \tag{17}$$

Let OS_t denote the set of true outliers in the dataset. The True Positive Rate (TPR) and the False Positive Rate (FPR) are respectively computed as

$$TPR(v) = \frac{|OS_d \bigcap OS_t|}{|OS_t|}, \tag{18}$$

$$FPR(v) = \frac{|OS_d - OS_t|}{|S - OS_t|}. \tag{19}$$

ROC is a popular evaluation index for anomaly detection, which has the advantage of being monotonic and easier to interpret. ROC plots $FPR(v)$ on the x-axis and $TPR(v)$ on the y-axis. The closer the ROC curve of a detection algorithm is to the upper left corner of the coordinate system, the higher its detection performance. However, it may be difficult to determine which algorithm is absolutely superior by ROC when the ROC curves of both algorithms are positioned at approximately the same level. For this reason, AUC is further proposed for evaluating the overall effectiveness of the algorithm. For $s_o \in OS_t, s_i \in S - OS_t = OS_t^c$, let $P(s_o, s_i)$ denote the probability that the abnormal–normal point pair (s_o, s_i) is correctly ranked (i.e., s_o appears before s_i) in the ranking of anomaly score. The AUC value is calculated by the average value of the probability $P(s_o, s_i)$ of all abnormal–normal point pairs, and its calculation formula is as follows.

$$\mathrm{AUC} = \frac{1}{|OS_t||OS_t^c|} \sum_{s_o \in OS_t, s_i \in OS_t^c} P(s_o, s_i), \tag{20}$$

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

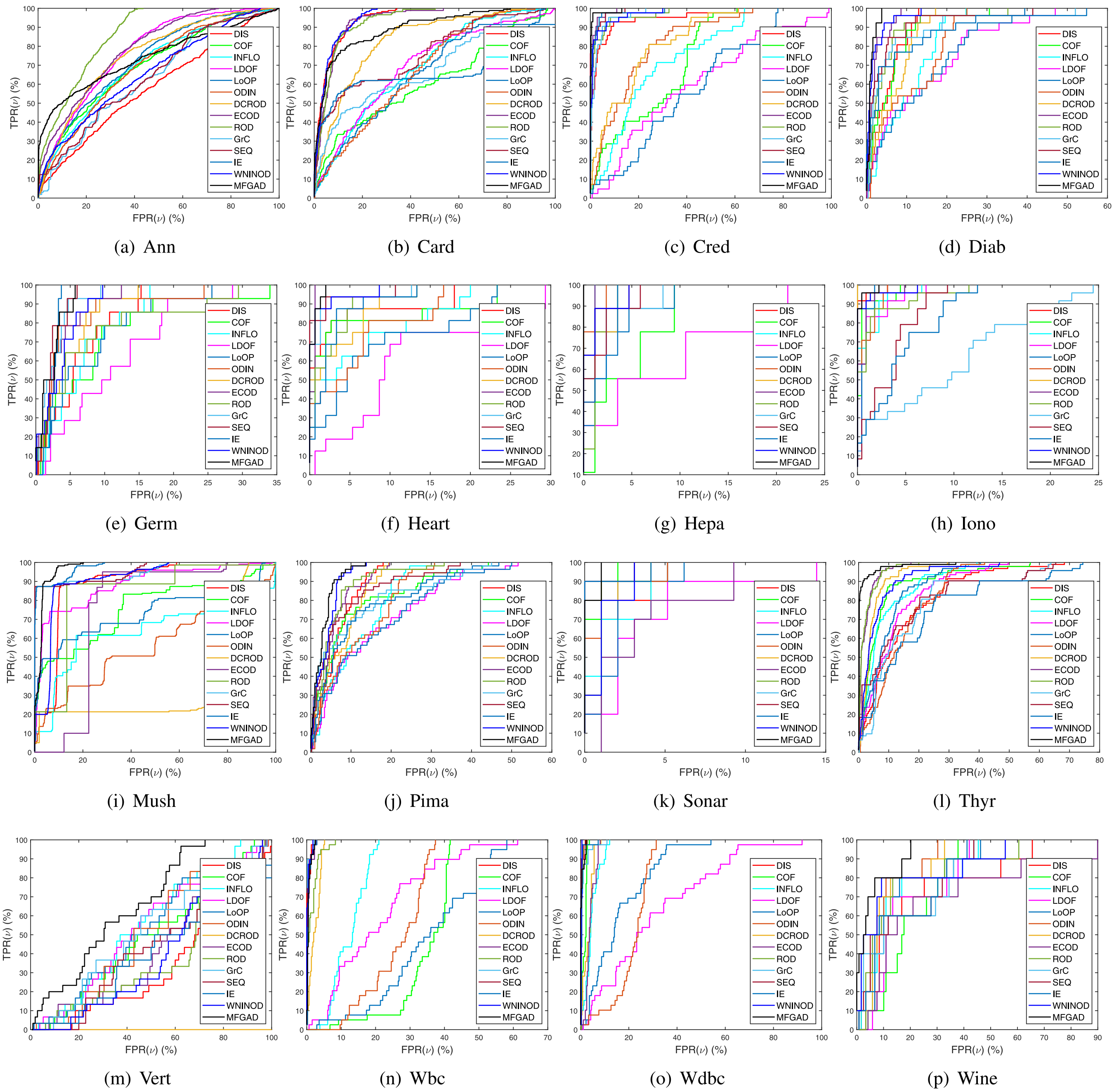


Fig. 1. Experimental comparison results on ROC.

where

$$P(s_o, s_i) = \begin{cases} 1, \text{if } Score(s_o) > Score(s_i); \\ \frac{1}{2}, \text{if } Score(s_o) = Score(s_i); \\ 0, \text{if } Score(s_o) < Score(s_i). \end{cases}$$

The value range of AUC is $[0, 1]$. The larger AUC value of an anomaly detection algorithm, the better its performance. In addition, the AUC value is an average on the probabilities of abnormal–normal point pairs and requires no additional parameters. Therefore, algorithm MFGAD only needs to determine the adjustment parameter δ of the fuzzy radius threshold ε.

5.2. Experimental results

First, we analyze the comparison results regarding ROC. The ROC curves of fourteen outlier detection algorithms are shown in Fig. 1. Among them, the black curves show the algorithms proposed in this paper. With Fig. 1, it is clear that the ROC curves of each comparison algorithm are monotonically undiminished. As mentioned earlier, the closer the ROC curve of a detection algorithm is to the upper left corner of the coordinate system, the higher its detection performance is.

Through Fig. 1, we can see that the ROC curves of the proposed algorithm MFGAD on some datasets are clearly closest to the upper left corner of the first quadrant, such as datasets Diab, Hepa, Pima, Thyr, Vert, and Wine. Herein, the ROC curves of algorithm MFGAD on dataset Hepa overlap exactly with the coordinates, which indicates that its exhibits full performance on datasets Hepa. In addition, the ROC curve of algorithm MFGAD is closer to the upper left corner of the first quadrant on some other datasets, such as datasets Cred, Heart, and Mush. Therefore, we can conclude that algorithm MFGAD exhibits superior performance in most cases.

Table 3
Experimental comparison results on AUC.

Datasets	DIS	COF	INFLO	LDOF	LoOP	ODIN	DCROD	ECOD	ROD	GrC	SEQ	IE	WNINOD	MFGAD
Ann	0.589	0.703	0.703	0.772	0.731	0.696	0.757	0.789	**0.860**	0.639	0.649	0.649	0.667	0.746
Card	0.946	0.620	0.698	0.698	0.667	0.686	0.834	0.935	0.932	0.690	0.766	0.672	**0.949**	0.897
Cred	0.954	0.746	0.760	0.616	0.598	0.832	0.841	0.990	0.974	0.975	0.990	0.986	0.979	**0.995**
Diab	0.952	0.942	0.902	0.863	0.870	0.898	0.935	0.979	0.954	0.940	0.957	0.930	0.981	**0.989**
Germ	0.938	0.918	0.932	0.889	0.925	0.952	0.955	0.966	0.927	0.974	0.976	0.976	0.965	**0.978**
Heart	0.970	0.957	0.945	0.880	0.923	0.949	0.970	0.996	0.975	0.988	0.991	0.975	0.992	**0.997**
Hepa	0.990	0.959	0.995	0.922	0.976	0.997	0.990	0.996	0.987	0.984	0.987	0.984	0.992	**1.000**
Iono	0.998	0.994	0.991	0.988	0.993	0.993	**1.000**	0.994	0.987	0.907	0.968	0.958	0.996	0.999
Mush	0.893	0.758	0.632	0.887	0.753	0.586	0.596	0.759	0.847	0.951	0.955	**0.979**	0.912	0.974
Pima	0.937	0.908	0.894	0.855	0.860	0.889	0.928	0.947	0.947	0.904	0.929	0.892	0.963	**0.971**
Sonar	0.984	0.996	0.984	0.964	0.981	0.994	0.989	0.966	0.990	0.997	0.997	0.996	0.989	**0.998**
Thyr	0.862	0.926	0.920	0.887	0.909	0.863	0.955	0.977	0.977	0.846	0.862	0.815	0.943	**0.989**
Vert	0.351	0.489	0.556	0.535	0.533	0.551	0.500	0.420	0.390	0.532	0.440	0.418	0.396	**0.667**
Wbc	**0.997**	0.654	0.875	0.792	0.645	0.734	0.976	0.995	0.985	0.996	0.996	0.995	**0.997**	0.994
Wdbc	0.995	0.984	0.954	0.702	0.839	0.786	0.968	0.959	0.995	0.990	**0.998**	0.996	0.996	0.996
Wine	0.797	0.798	0.850	0.816	0.855	0.887	0.885	0.733	0.850	0.827	0.833	0.829	0.873	**0.945**
Average	0.885	0.834	0.849	0.817	0.816	0.831	0.880	0.900	0.911	0.884	0.893	0.878	0.912	**0.946**

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

Second, the results of algorithm MFGAD and other comparison algorithms regarding AUC are shown in Table 3, where the bolded part shows the best results among all fourteen algorithms.

Based on these sixteen datasets, Table 3 can better reflect the advantages of the proposed detection algorithm. Obviously, algorithm MFGAD achieves better AUC values in most cases. For example, the AUC value of algorithm MFGAD on dataset Cred is 0.995. However, the AUC values of algorithms DIS, COF, INFLO, LDOF, LoOP, ODIN, DCROD, ECOD, ROD, GrC, SEQ, IE, and WNINOD are only 0.954, 0.746, 0.760, 0.616, 0.598, 0.832, 0.841, 0.990, 0.974, 0.975, 0.990, 0.986, and 0.979, respectively, which are all smaller than that of algorithm MFGAD. From a statistical point of view, algorithm MFGAD achieves the best results on 10 datasets; the remaining algorithms DIS, COF, INFLO, LDOF, LoOP, ODIN, DCROD, ECOD, ROD, GrC, SEQ, IE, and WNINOD achieve the best results only on 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, and 2, respectively. That is to say, algorithm MFGAD always achieves better performance in both individual and overall statistical situations. The final average of the AUC values also can better reveal and validate comparative performance. Algorithms DIS, COF, INFLO, LDOF, LoOP, ODIN, DCROD, ECOD, ROD, GrC, SEQ, IE, WNINOD, and MFGAD correspond to 0.885, 0.834, 0.849, 0.817, 0.816, 0.831, 0.880, 0.900, 0.911, 0.884, 0.893, 0.878, 0.912, and 0.946, respectively. Among them, algorithm MFGAD obtains the best value of 0.946, which is significantly larger than those of other algorithms.

In addition, these datasets include nominal data, numerical data, and hybrid attribute data. In conclusion, through comparative analyses, algorithm MFGAD can effectively process three types of attribute data.

Third, Friedman test [55] and Nemenyi post-hoc test [56] are used to evaluate the statistical significance of the results. Before using Friedman test, AUC value of each algorithm on all datasets is sorted from low to high, and the sequence number is assigned $(1, 2, \ldots)$. Among them, if AUC value of the two algorithms is the same, the ordinal values are equally divided. Then, Friedman test is used to determine whether these algorithms have the same performance. Suppose we compare M algorithms on N datasets, and let r_i denote the average ordinal value of the ith algorithm, then Friedman test is calculated as follows.

$$\tau_F = \frac{(N-1)\tau_{\chi^2}}{N(M-1)-\tau_{\chi^2}} \text{ and }$$
$$\tau_{\chi^2} = \frac{12N}{M(M+1)}\left(\sum_{i=1}^{M} r_i^2 - \frac{M(M+1)^2}{4}\right). \tag{21}$$

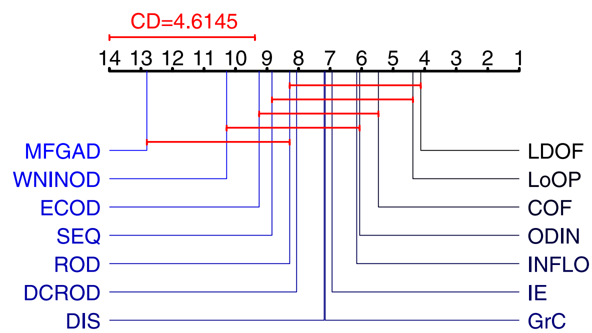


Fig. 2. Nemenyi test figure on AUC.

τ_F obeys the F distribution with $(M-1)$ and $(M-1)(N-1)$ degrees of freedom. If the null hypothesis of "all algorithms have the same performance" is rejected, it means that the performance of the algorithms is significantly different. At this time, a post-hoc test needs to be used to further distinguish these feature selection algorithms. Nemenyi post-hoc test is commonly used. In Nemenyi test, the critical difference (CD) of the average ordinal value is calculated by the following formula.

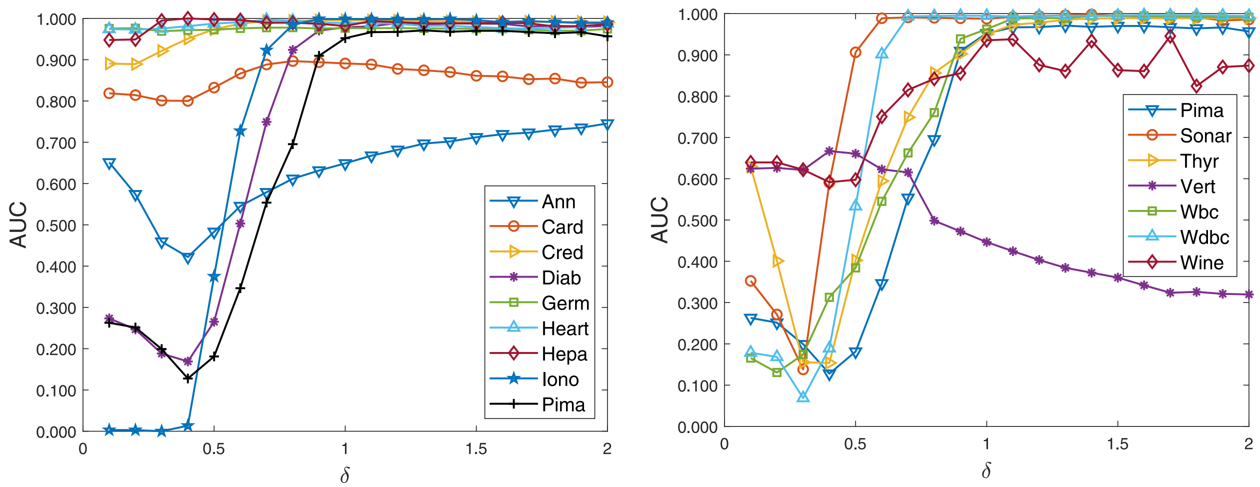
$$CD_\alpha = q_\alpha \sqrt{\frac{M(M+1)}{6N}}, \tag{22}$$

where q_α is the critical value of Tukey's distribution, which can be found in [56].

Nemenyi test figure can be used to more intuitively represent the significant differences between the two algorithms [43]. In Nemenyi test figure, for each algorithm, a dot is used to show its average ordinal value, and a horizontal line segment with the dot as the center is used to indicate the size of CD. If a group of algorithms is connected by horizontal line segments, then it means that there is no significant difference between this group of algorithms.

Specifically, we can get $M = 14$ and $N = 16$, the τ_F distribution has 13 and 195 degrees of freedom. According to Friedman test, when $\alpha = 0.1$, the value of $\tau_F = 6.8665$ is greater than the critical value 1.5585. Therefore, the null hypothesis that "all algorithms have the same performance" is rejected. It shows that the performance of all outlier detection algorithms is significantly different. At this time, a post-hoc test needs to be used to further distinguish them.

For significance level $\alpha = 0.1$, the corresponding critical distance $CD_{0.1} = 4.6145$ can be obtained. Finally, Nemenyi test figure on AUC is shown in Fig. 2. From Fig. 2, we can see that algorithm MFGAD is statistically significantly different from most other algorithms. For example, it can be seen from Fig. 2 that algorithm MFGAD is not connected to algorithms DIS, COF, INFLO, LDOF, LoOP, ODIN, DCROD,



(a) Datasets 1-8
(b) Datasets 10-16

Fig. 3. Variation curve of AUC with parameter δ.

GrC, and IE with horizontal line segments, which indicates that MFGAD is statistically significantly different from these algorithms. However, there is no consistent evidence to indicate the statistical differences from algorithms WNINOD, ECOD, SEQ, and ROD.

5.3. Parameter sensitivity analyses

In algorithm MFGAD, the parameter δ plays an important role in anomaly detection because it determines the size of the fuzzy radius ε. Therefore, we further focus on parameter sensitivity analysis to reveal the relationship between δ changes and detection results. Through experiments only on algorithm MFGAD, Fig. 3 gives the AUC curve with δ. Note that the detection of nominal data (e.g., dataset Mush) does not involve parameters.

It can be seen from Fig. 3 that with the increase of the parameter δ, the AUC curve of algorithm MFGAD on most datasets first increases rapidly and then tends to be stable. However, with the increase of the parameter δ, the AUC curve of algorithm MFGAD on some datasets first increases rapidly and then decreases, such as dataset Heart. At the same time, for different datasets, algorithm MFGAD can obtain the optimal AUC value under multiple δ, such as datasets Diab, Iono, Sonar, and Cred.

It can be seen from the above analysis results that almost all datasets are sensitive to changes in the parameter δ, and different situations may cause different results. Therefore, how to scientifically determine the value or range of δ is a fundamental problem to improve the performance of algorithm MFGAD.

6. Conclusions

In this paper, a multi-fuzzy granules-based anomaly detection model is constructed by using the fuzzy rough computing model. The proposed model not only takes into account the advantages of fuzzy rough computing models that can effectively deal with fuzzy or uncertainty data, but also incorporates the idea of multi-granularity to understand and analyze the problem of anomaly detection. In the proposed model, a hybrid metric is employed to compute fuzzy relations. As a result, the proposed model is suitable for nominal, numerical, and hybrid attribute data. Finally, a corresponding MFGAD algorithm is designed to detect anomalies. The comparative experimental results on public data demonstrate the effectiveness and applicability of the proposed algorithm. The results of the parameter sensitivity analysis show that the proposed algorithm has a high sensitivity to the parameters. Therefore, in future work, it is necessary to investigate how to determine the optimal parameter values adaptively. In addition, contextual or conditional outlier detection techniques will be further explored depending on different application scenarios.

CRediT authorship contribution statement

Zhong Yuan: Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Hongmei Chen:** Project administration, Validation, Funding acquisition. **Chuan Luo:** Project administration, Validation, Funding acquisition. **Dezhong Peng:** Project administration, Validation, Funding acquisition.

Declaration of competing interest

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

Data availability

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

Acknowledgments

The authors thank both the editors and reviewers for their valuable suggestions, which substantially improve this paper. This work was supported by the National Natural Science Foundation of China (61976182 and 62076171), Sichuan Science and Technology Planning Project (2023YFQ0020, 2021YFG0317, 2022YFQ0014, 2022YFH0021, 2021YFS0389, and 2023YFG0033), and the Fundamental Research Funds for the Central Universities, China (YJ202245).

References

[1] B. Wang, Z.Z. Mao, A dynamic ensemble outlier detection model based on an adaptive k-nearest neighbor rule, Inf. Fusion 63 (2020) 30–40.
[2] F. Jiang, G.Z. Liu, J.W. Du, Y.F. Sui, Initialization of K-modes clustering using outlier detection techniques, Inform. Sci. 332 (2016) 167–183.
[3] Z. Zhang, Q.L. Feng, J.Y. Huang, Y.T. Guo, J.H. Xu, J.X. Wang, A local search algorithm for k-means with outliers, Neurocomputing 450 (2021) 230–241.
[4] S. Panigrahi, A. Kundu, S. Sural, A.K. Majumdar, Credit card fraud detection: A fusion approach using Dempster–Shafer theory and Bayesian learning, Inf. Fusion 10 (4) (2009) 354–363.
[5] B. Wang, Z.Z. Mao, Outlier detection based on a dynamic ensemble model: Applied to process monitoring, Inf. Fusion 51 (2019) 244–258.

[6] H.H. Bosman, G. Iacca, A. Tejada, H.J. Wörtche, A. Liotta, Spatial anomaly detection in sensor networks using neighborhood information, Inf. Fusion 33 (2017) 41–56.
[7] J.T. Yao, A.V. Vasilakos, W. Pedrycz, Granular computing: perspectives and challenges, IEEE Trans. Cybern. 43 (6) (2013) 1977–1989.
[8] A. Tan, J.Y. Liang, W.-Z. Wu, J. Zhang, L. Sun, C. Chen, Fuzzy rough discrimination and label weighting for multi-label feature selection, Neurocomputing 465 (2021) 128–140.
[9] Y.W. Li, Y.J. Lin, J.H. Liu, W. Weng, Z.K. Shi, S.X. Wu, Feature selection for multi-label learning based on kernelized fuzzy rough sets, Neurocomputing 318 (2018) 271–286.
[10] 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 Trans. Fuzzy Syst. 30 (9) (2022) 3395–3409.
[11] H.-X. Zhao, H.-J. Xing, X.-Z. Wang, Two-stage dimensionality reduction approach based on 2DLDA and fuzzy rough sets technique, Neurocomputing 74 (17) (2011) 3722–3727.
[12] B.B. Sang, H.M. Chen, L. Yang, T.R. Li, W.H. Xu, Incremental feature selection using a conditional entropy based on fuzzy dominance neighborhood rough sets, IEEE Trans. Fuzzy Syst. 30 (6) (2022) 1683–1697.
[13] B.Z. Sun, X.M. Zhou, N.N. Lin, Diversified binary relation-based fuzzy multigranulation rough set over two universes and application to multiple attribute group decision making, Inf. Fusion 55 (2020) 91–104.
[14] B. Cao, J.W. Zhao, Z.H. Lv, Y. Gu, P. Yang, S.K. Halgamuge, Multiobjective evolution of fuzzy rough neural network via distributed parallelism for stock prediction, IEEE Trans. Fuzzy Syst. 28 (5) (2020) 939–952.
[15] X.D. Yue, J. Zhou, Y.Y. Yao, D.Q. Miao, Shadowed neighborhoods based on fuzzy rough transformation for three-way classification, IEEE Trans. Fuzzy Syst. 28 (5) (2020) 978–991.
[16] Z. Yuan, H.M. Chen, T.R. Li, B.B. Sang, S. Wang, Outlier detection based on fuzzy rough granules in mixed attribute data, IEEE Trans. Cybern. 52 (8) (2022) 8399–8412.
[17] Z. Yuan, H.M. Chen, T.R. Li, J. Liu, S. Wang, Fuzzy information entropy-based adaptive approach for hybrid feature outlier detection, Fuzzy Sets and Systems 421 (2021) 1–28.
[18] Z. Yuan, B.Y. Chen, J. Liu, H.M. Chen, D.Z. Peng, P.L. Li, Anomaly detection based on weighted fuzzy-rough density, Appl. Soft Comput. 134 (2023) 109995.
[19] P.F. Zhang, T.R. Li, Z. Yuan, C. Luo, G.Q. Wang, J. Liu, S.D. Du, A data-level fusion model for unsupervised attribute selection in multi-source homogeneous data, Inf. Fusion 80 (2022) 87–103.
[20] P. Hu, H.Y. Zhu, J. Lin, D.Z. Peng, Y.-P. Zhao, X. Peng, Unsupervised contrastive cross-modal hashing, IEEE Trans. Pattern Anal. Mach. Intell. http://dx.doi.org/10.1109/TPAMI.2022.3177356.
[21] Y.H. Qian, J.Y. Liang, Y.Y. Yao, C.Y. Dang, MGRS: A multi-granulation rough set, Inform. Sci. 180 (6) (2010) 949–970.
[22] P.F. Zhang, T.R. Li, C. Luo, G.Q. Wang, AMG-DTRS: Adaptive multi-granulation decision-theoretic rough sets, Internat. J. Approx. Reason. 140 (2022) 7–30.
[23] J.M. Zhan, B.Z. Sun, J.C.R. Alcantud, Covering based multigranulation (I, T)-fuzzy rough set models and applications in multi-attribute group decision-making, Inform. Sci. 476 (2019) 290–318.
[24] W.P. Ding, W. Pedrycz, I. Triguero, Z.H. Cao, C.-T. Lin, Multigranulation supertrust model for attribute reduction, IEEE Trans. Fuzzy Syst. 29 (6) (2020) 1395–1408.
[25] W. Wei, J.Y. Liang, Information fusion in rough set theory: An overview, Inf. Fusion 48 (2019) 107–118.
[26] P.F. Zhang, T.R. Li, G.Q. Wang, C. Luo, H.M. Chen, J.B. Zhang, D.X. Wang, Z. Yu, Multi-source information fusion based on rough set theory: A review, Inf. Fusion 68 (2021) 85–117.
[27] 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.
[28] F. Jiang, Y.F. Sui, C.G. Cao, A rough set approach to outlier detection, Int. J. Gen. Syst. 37 (5) (2008) 519–536.
[29] F. Jiang, Y.F. Sui, C.G. Cao, Some issues about outlier detection in rough set theory, Expert Syst. Appl. 36 (3) (2009) 4680–4687.
[30] F. Jiang, Y.F. Sui, C.G. Cao, An information entropy-based approach to outlier detection in rough sets, Expert Syst. Appl. 37 (9) (2010) 6338–6344.
[31] F. Jiang, Y.F. Sui, C.G. Cao, A hybrid approach to outlier detection based on boundary region, Pattern Recognit. Lett. 32 (14) (2011) 1860–1870.
[32] Y.M. Chen, D.Q. Miao, R.Z. Wang, Outlier detection based on granular computing, in: International Conference on Rough Sets and Current Trends in Computing, Springer, 2008, pp. 283–292.
[33] A. Albanese, S.K. Pal, A. Petrosino, Rough sets, kernel set, and spatiotemporal outlier detection, IEEE Trans. Knowl. Data Eng. 26 (1) (2012) 194–207.
[34] F. Jiang, Y.M. Chen, Outlier detection based on granular computing and rough set theory, Appl. Intell. 42 (2) (2015) 303–322.
[35] 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, Decis. Support Syst. 75 (2015) 63–75.
[36] F. Jiang, H.B. Zhao, J.W. Du, Y. Xue, Y.J. Peng, Outlier detection based on approximation accuracy entropy, Int. J. Mach. Learn. Cybern. 10 (9) (2019) 2483–2499.
[37] M. Singh, R. Pamula, An outlier detection approach in large-scale data stream using rough set, Neural Comput. Appl. 32 (13) (2020) 9113–9127.
[38] Y.M. Chen, D.Q. Miao, H.Y. Zhang, Neighborhood outlier detection, Expert Syst. Appl. 37 (12) (2010) 8745–8749.
[39] Z. Yuan, X.Y. Zhang, S. Feng, Hybrid data-driven outlier detection based on neighborhood information entropy and its developmental measures, Expert Syst. Appl. 112 (2018) 243–257.
[40] 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 Trans. Syst. Man Cybern. Syst. 52 (8) (2022) 5175–5187.
[41] Y. Wang, Y.P. Li, Outlier detection based on weighted neighbourhood information network for mixed-valued datasets, Inform. Sci. 564 (2021) 396–415.
[42] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[43] Z. Yuan, H.M. Chen, P. Xie, P.F. Zhang, J. Liu, T.R. Li, Attribute reduction methods in fuzzy rough set theory: An overview, comparative experiments, and new directions, Appl. Soft Comput. 107 (2021) 107353.
[44] Q.H. Hu, D.R. Yu, Z.X. Xie, J.F. Liu, Fuzzy probabilistic approximation spaces and their information measures, IEEE Trans. Fuzzy Syst. 14 (2) (2006) 191–201.
[45] E.M. Knorr, R.T. Ng, V. Tucakov, Distance-based outliers: algorithms and applications, VLDB J. 8 (3) (2000) 237–253.
[46] 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.
[47] 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.
[48] 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.
[49] 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.
[50] V. Hautamaki, I. Karkkainen, P. Franti, Outlier detection using k-nearest neighbour graph, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004, Vol. 3, ICPR 2004, IEEE, 2004, pp. 430–433.
[51] K.S. Li, X. Gao, S.Y. Fu, X.P. Diao, P. Ye, B. Xue, J.H. Yu, Z.J. Huang, Robust outlier detection based on the changing rate of directed density ratio, Expert Syst. Appl. 207 (2022) 117988.
[52] Z. Li, Y. Zhao, X.Y. Hu, N. Botta, C. Ionescu, G. Chen, ECOD: Unsupervised outlier detection using empirical cumulative distribution functions, IEEE Trans. Knowl. Data Eng. http://dx.doi.org/10.1109/TKDE.2022.3159580.
[53] Y. Almardeny, N. Boujnah, F. Cleary, A novel outlier detection method for multivariate data, IEEE Trans. Knowl. Data Eng. 34 (9) (2022) 4052–4062.
[54] G.O. Campos, A. Zimek, J. Sander, R.J. Campello, B. Micenková, E. Schubert, I. Assent, M.E. Houle, On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study, Data Min. Knowl. Discov. 30 (4) (2016) 891–927.
[55] M. Friedman, A comparison of alternative tests of significance for the problem of m rankings, Ann. Math. Stat. 11 (1) (1940) 86–92.
[56] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (Jan) (2006) 1–30.