Information Fusion 103 (2024) 102133

Contents lists available at ScienceDirect

Information Fusion

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

Full length article

Fusing multi-scale fuzzy information to detect outliers

Baiyang Chen [a,1], Yongxiang Li [a,1], Dezhong Peng [a,b], Hongmei Chen [c], Zhong Yuan [a,*]

[a] College of Computer Science, Sichuan University, Chengdu, 610065, China
[b] Chengdu Ruibei Yingte Information Technology Co., Ltd, Chengdu, 610054, China
[c] School of Computing and Artificial Intelligence, Southwest Jiaotong University, Chengdu, 611756, China

ARTICLE INFO

Keywords:
Outlier detection
Information fusion
Fuzzy rough sets
Multi-scale fuzzy granules
Fuzzy approximation

ABSTRACT

Outlier detection aims to find objects that behave differently from the majority of the data. Existing unsupervised approaches often process data with a single scale, which may not capture the multi-scale nature of the data. In this paper, we propose a novel information fusion model based on multi-scale fuzzy granules and an unsupervised outlier detection algorithm with the fuzzy rough set theory. First, a multi-scale information fusion model is formulated based on fuzzy granules. Then we employ fuzzy approximations to define the outlier factor of multi-scale fuzzy granules centered at each data point. Finally, the outlier score is calculated by aggregating the outlier factors of a set of multi-scale fuzzy granules. Experimental results demonstrate that the proposed method is comparable with or better than the leading outlier detection methods. The codes and datasets are publicly available online at https://github.com/ChenBaiyang/MFIOD.

1. Introduction

Outlier detection (OD), also referred to as anomaly detection or novelty detection, aims to identify abnormal data points whose behavior is significantly different from the majority of the data. Detecting outliers is usually a necessary step in knowledge discovery and data mining, as it eliminates incorrect or inaccurate observations that might otherwise distort our conclusions. Hence, OD has numerous applications, such as fraud detection [1], network intrusion detection [2], industry control [3], medical anomaly detection [4], etc.

Fuzzy Rough Sets (FRS) [5], which unifies rough sets and fuzzy sets, is a Granular computing (GrC) model for processing data with uncertainty or imprecision. The fundamental concept of FRS is to utilize various fuzzy membership functions to capture the boundaries of a concept in uncertain or imprecise data, and then use fuzzy approximations to facilitate decision-making processes. This allows us to classify objects even if they do not exactly fit the criteria for a certain class, and to handle heterogeneous data without converting the data types. FRS has been successfully applied to various domains including feature selection [6–8] and outlier detection [9–11] in heterogeneous data over the past years. However, these detection methods are only applicable with a single-scale view, which cannot represent the same object from different scale levels. Generally, an attribute of the same instance may have various values when it is measured in multiple units, for instance, in the geographical information systems and remote sensing data recorded at multiple scales [12]. Wu and Leung [13] are the first to formalize a GrC model for dealing with multi-scale data that can be measured at various scale levels. So far, this model has been applied to many research fields including feature selection [14,15], rule acquisition [16], decision analysis [17], etc. If multi-scale information is integrated properly, the performance and reliability of the detection algorithm will have the potential to improve. To this end, this paper formulates a multi-scale FRS model and proposes to fuse *M*ulti-scale *F*uzzy *I*nformation for unsupervised *O*utlier *D*etection (MFIOD) in heterogeneous data.

The idea behind our method can be summarized as follows. An outlier in data can be defined as an object that deviates significantly from the majority of samples or exhibits unique characteristics. From the perspective of FRS, this uniqueness manifests as the difficulty of outliers to be approximated by most normal objects. Therefore, this paper employs the notion of fuzzy approximation from FRS to characterize outliers in data. Specifically, we first obtain multi-scale data representations by fuzzy information granulation, and then propose an information fusion model to construct multi-scale fuzzy granules (MFG). Second, we use a set of attribute subsets to construct a series of MFG for each data point, then select some other granules to approximate these MFG, and quantify the quality of the approximations by a metric from the FRS theory. Finally, if an object has a poor average approximation quality

* Corresponding author.
E-mail addresses: farstars@qq.com (B. Chen), 838194547@qq.com (Y. Li), pengdz@scu.edu.cn (D. Peng), hmchen@swjtu.edu.cn (H. Chen), yuanzhong@scu.edu.cn (Z. Yuan).
[1] Equal contribution.

https://doi.org/10.1016/j.inffus.2023.102133
Received 19 May 2023; Received in revised form 4 November 2023; Accepted 6 November 2023
Available online 8 November 2023

Table 1
The advantages and limitations of unsupervised outlier detectors.

Categories	Methods	Advantages	Limitations
Statistical	GMM [18]	+ Insensitive to data distribution	- High computational complexity
	HBOS [19]	+ Handle data streams	- Sensitive to bin sizes - Cannot detect local outliers
	ECOD [20]	+ Low computational cost + Parameter free	- Sensitive to probabilistic distribution
Distance-based	IForest [21]	+ Insensitive to data distribution + Handle high-dimensional data + Low computational cost	- Sensitive to parameter settings
	LDOF [22]	+ Effective at local anomalies + Insensitive to data distribution	- Sensitive to parameter settings - High computational complexity
	FastSolvingSet [23]	+ High computational efficiency	- Sensitive to parameter settings
Density-based	LOF [24]	+ Effective for local anomalies + Insensitive to data distribution	- Sensitive to the parameter k
	LoOP [25]	+ Probabilistic interpretation of outliers	- Parameter difficult to select
	WDOD [26]	+ High computational efficiency + Handle categorical data	- Sensitive to parameter settings
	NOF [27]	+ Parameter free + High computational efficiency	- Lower performance
	Hu et al. [28]	+ Insensitive to data distribution + Effective for local anomalies	- Sensitive to parameter settings
	DCROD [29]	+ Insensitive to data distribution	- Sensitive to the parameter k
Clustering-based	FindCBLOF [30]	+ Interpretable results + Robust to data distributions	- Sensitive to initialization of clustering
	CBOF [31]	+ Detect both individual and collective outliers	- Too many parameters to tune
	ROCF [32]	+ Detect both individual and collective outliers	- Lower performance
GrC-based	GrC [33]	+ Handle uncertain or imprecise data + Robust to data distributions	- Require discretization for nominal data - High computational cost
	IE [34]	+ Handle uncertain or imprecise data	- Require discretization for nominal data - High computational cost
	ODGrCR [35]	+ Handle uncertain or imprecise data + Robust to data distributions	- Require discretization for nominal data
	Singh & Pamula [36]	+ Handle large-scale data streams + High computational efficiency	- Require discretization for nominal data
	NED [37]	+ Adapt to numerical and nominal data	- High computational cost
	FRGOD [38]	+ Adapt to numerical and nominal data	- High computational cost
	WFRDA [39]	+ Low computational cost + Adapt to numerical and nominal data	- Sensitive to parameter settings
	FKMOD [11]	+ Adapt to numerical and nominal data	- Sensitive to parameter settings

across a set of MFG, then this object possesses greater uniqueness, and is more likely to be an outlier. The contributions of this paper include:

• This paper proposes a new FRS model that applies multi-scale knowledge representation to outlier detection and achieves satisfactory results.

• A novel information fusion model based on multi-scale fuzzy granules is formalized, and one may analyze and integrate data from different scales.

• Extensive experiments demonstrate that the proposed algorithm enjoys higher precision over some state-of-the-art models on 15 public datasets.

The paper is organized as follows. Section 2 reviews the related works on outlier detection, and highlights the main challenges and limitations of the existing methods. Section 3 introduces some preliminaries and definitions that are necessary for understanding our approach. Section 4 describes our methodology in detail. Section 5 reports the experiments and analysis. Section 6 concludes the paper and discusses some future directions for research.

2. Related works

Unsupervised outlier detection methods can be generally categorized into five classes including statistics-based, distance-based, density-based, cluster-based, as well as GrC-based approaches. It is notable that some methods may belong to more than one group or may combine elements from different groups. Table 1 summarizes the advantages and limitations of the detection methods discussed in this section.

Statistical-based Methods assume that outliers exhibit extreme values or deviate significantly from the expected statistical properties (e.g., mean/standard deviation [18], histogram [19]) of the data. These methods are relatively easy to understand and implement, but generally sensitive to the distribution of data. For example, Li et al. [20] designed a non-parametric statistical model based on the empirical cumulative distribution of data for outlier detection (ECOD).

Distance-based Methods attempt to find anomalies by directly comparing the distance of an arbitrary pair of objects and any objects far away from the rest of data are regarded as outliers. This approach can handle data with multiple modes (multimodal data) and is usually free of data distribution. A remarkable detection method by Liu et al. [21] was IForest which isolates objects with binary trees and identifies instances far from the root as outliers. Zhang et al. [22] presented a local distance-based method (LDOF) to determine the extent to which an object deviates from its dispersed neighborhood. Angiulli et al. [23] put forward inverse triangular inequalities to improve the computational efficiency of pairwise distance calculations. However, these methods generally require calculating pairwise distances of all objects, thus suffering from low computation efficiency.

Density-based Methods adopt the assumption that outliers are located in regions of low data density and identify instances with a significantly lower density as potential outliers. These methods can effectively handle datasets with varying data densities and are robust to data distributions, making them suitable for data with irregular shapes and varying local densities.For example, Breunig et al. [24] introduced the concept of local outlier factor (LOF) for outlier detection. Kriegel et al. [25] combined the idea of local outlier score with probability statistics and presented a local outlier probability detection model (LoOP). However, a large number of computational resources are required to compute the LOF for each object. To solve this problem, Zhao et al. [26] investigated the density and uncertainty of objects and formalized an effective detection algorithm (WDOD). In addition, these methods generally require extensive hyperparameter tuning, which greatly affects performance and efficiency. In view of this, Huang et al. [27] formulated a non-parametric outlier detection algorithm Natural Outlier Factor (NOF) by exploiting natural neighbors. Hu et al. [28] designed a local kernel to calculate the local densities of objects and used the weighted neighborhood density estimation to enhance the model's robustness. Recently, Li et al. [29] combined kernel density estimation together with a neighborhood set for computing the local density and proposed an effective anomaly detection algorithm (DCROD).

Clustering-based methods suppose that outliers do not belong to any cluster or belong to small, sparse clusters. This type of method is often interpretable and robust to data distributions, but may be sensitive to the initial conditions of clustering. One such method by He et al. [30] was the Cluster-Based Local Outlier Factor (CBLOF) that detects outliers based on cluster assignments. Similarly, Duan et al. [31] attempt to find both individual and collective outliers (CBOF) simultaneously. Following CBLOF, Huang et al. [32] proposed the Relative Outlier Cluster Factor (ROCF) to specify a cluster size gap to divide normal and abnormal clusters.

GrC-based Methods aim to represent and analyze data in a way that allows for effective outlier detection by considering various levels of granularity or abstraction. This approach can adapt to diverse data types and structures, making them suitable for various applications, including feature selection [7], outlier detection [33] and decision-making [40]. However, they usually come with computational challenges and may require careful parameter tuning. For example, Chen et al. [33] proposed a Granular Computing-based anomaly detection algorithm (GrC) that used information granules to express the degree of outliers. Jiang et al. [34] employed the notion of information entropy to measure the uncertainty of data in their outlier detection model (IE) in 2010, and they further put forward an extension anomaly detection algorithm (ODGrCR) by rough approximation accuracy [35] in 2015. Singh and Pamula [36] applied Rough Sets in a large-scale data stream for anomaly detection. However, the above-mentioned methods construct OD models based on equivalence classes. Therefore they generally do not apply to numerical values, and require discretization operation in pre-processing steps, which bring much information loss. To this end, some researchers make efforts to extend the classical Rough Set theory to deal with numerical data. Concretely, Chen et al. [37] took the neighborhood rough sets as the basic framework which can avoid extra discretization (NED). Yuan et al. further put forward several solutions, such as fuzzy rough granules [38] and fuzzy-rough density [39] for outlier detection in both numerical and nominal data. Lately, Zhang et al. [11] designed fuzzy knowledge measure (FKM) to describe outliers (FKMOD), and proposed a two-stage detection model that includes fusion of multiple information sources.

3. Preliminaries

This section reviews some basic concepts and definitions in FRS and multi-scale information systems. We use the nomenclature in Table 2 to denote the notations used in this paper.

Table 2
Nomenclature.

Notation	Description
U	Set of objects
A	Set of attributes
B, P	Subsets of attributes $B, P \subseteq A$
x, y	Objects $x, y \in U$
a	An attribute, $a \in A$
$\mathcal{X}$	A fuzzy set
$\mathcal{R}$	A fuzzy relation
$M(\mathcal{R})$	Matrix of fuzzy relation $\mathcal{R}$
r_{ij}	(i, j)-th element of a fuzzy relation matrix
$G_U(\mathcal{R})$	Fuzzy granular structure of U generated by $\mathcal{R}$
$[x_i]_B$	Fuzzy granule centered at x_i generated by B
Λ	Set of scales in data
λ	A data scale, $\lambda \in \Lambda$
$\underline{\mathcal{R}_B}\mathcal{X}$	Lower approximation of $\mathcal{X}$
$\overline{\mathcal{R}_B}\mathcal{X}$	Upper approximation of $\mathcal{X}$
$\alpha_P(\cdot)$	Fuzzy approximation accuracy by P

3.1. Fuzzy rough sets

In this part, we briefly review some basic concepts of FRS including information systems, fuzzy sets, fuzzy relations, fuzzy granules, and fuzzy approximations. In applications of FRS, information is usually stored in an attribute-value table (also called an information system, information table, etc.), where every row indicates an object (instance), and every column denotes an attribute of each object.

Definition 1. An information system (IS) is a tuple (U, A), where $U = \{x_1, x_2, \ldots, x_n\}$ is the set of objects, also referred to as the universe of discourse, and $A = \{a_1, a_2, \ldots, a_m\}$ is the set of attributes that every object has.

Definition 2. Let $U = \{x_1, x_2, \ldots, x_n\}$ be the universe of discourse. If $\mathcal{X}$ is a map from U to $[0, 1]$, then $\mathcal{X}$ is a fuzzy set on U, i.e. $\mathcal{X} : U \to [0, 1]$.

$\forall x_i \in U$, $\mathcal{X}(x_i)$ is called as the membership of x_i for $\mathcal{X}$, or the membership function of $\mathcal{X}$. The fuzzy set is often denoted by $\mathcal{X} = (\mathcal{X}(x_1), \mathcal{X}(x_2), \ldots, \mathcal{X}(x_n))$.

Definition 3. Given a set of objects U, a fuzzy relation $\mathcal{R}$ on U is defined as a family of fuzzy sets $\mathcal{R} : U \times U \to [0, 1]$.

Some widely used operations of fuzzy relations are defined as follows.

(1) $\mathcal{R}_1 = \mathcal{R}_2 \Leftrightarrow \forall (x, y) \in U \times U, \mathcal{R}_1(x, y) = \mathcal{R}_2(x, y)$.
(2) $\mathcal{R}_1 \subseteq \mathcal{R}_2 \Leftrightarrow \forall (x, y) \in U \times U, \mathcal{R}_1(x, y) \leq \mathcal{R}_2(x, y)$.
(3) $(\mathcal{R}_1 \cup \mathcal{R}_2)(x, y) = \mathcal{R}_1(x, y) \vee \mathcal{R}_2(x, y) = \max\{\mathcal{R}_1(x, y), \mathcal{R}_2(x, y)\}$.
(4) $(\mathcal{R}_1 \cap \mathcal{R}_2)(x, y) = \mathcal{R}_1(x, y) \wedge \mathcal{R}_2(x, y) = \min\{\mathcal{R}_1(x, y), \mathcal{R}_2(x, y)\}$.

$\forall (x_i, x_j) \in U \times U$, the fuzzy degree $\mathcal{R}(x_i, x_j)$ expresses the degree to which x_i has a relation $\mathcal{R}$ with x_j. A fuzzy relation $\mathcal{R}$ on U is usually denoted by a fuzzy relation matrix $M(\mathcal{R}) = (r_{ij})_{n \times n}$, where $r_{ij} = \mathcal{R}(x_i, x_j)$. $\forall x_1, x_2, x_3 \in U$, a fuzzy relation $\mathcal{R}$ holds the following properties:

(1) $\mathcal{R}(x_1, x_1) = 1 \Leftrightarrow \mathcal{R}$ is reflexive.
(2) $\mathcal{R}(x_1, x_2) = \mathcal{R}(x_2, x_1) \Leftrightarrow \mathcal{R}$ is symmetric.
(3) $\mathcal{R}(x_1, x_2) \geq \min(\mathcal{R}(x_1, x_3), \mathcal{R}(x_3, x_2)) \Leftrightarrow \mathcal{R}$ is transitive.

A fuzzy relation has different connotations when the different properties mentioned above are achieved. Concretely, if $\mathcal{R}$ satisfies property 1 and property 2, $\mathcal{R}$ is called a fuzzy similarity (tolerance) relation. If $\mathcal{R}$ conforms to all the aforementioned properties, then $\mathcal{R}$ is also called a fuzzy equivalence relation.

Let $B \subseteq A$, a fuzzy similarity relation $\mathcal{R}_B$ on U can be derived from B. There are several means for calculating the membership degree of

$\mathcal{R}_B$, among which the conjunction is most widely used [41], which is computed by

$$\mathcal{R}_B(x_i, x_j) = \min_{a_k \in B} \mathcal{R}_{a_k}(x_i, x_j). \tag{1}$$

Granulation is a central idea of Granular computing. Information granulation involves the procedure of partitioning all objects into granules, where a granule is a collection or cluster of data elements that are arranged together due to their indistinguishability, similarity, proximity, etc. By working with granules, it allows for more flexibility and adaptability in handling diverse types of data and knowledge. In classic rough sets, objects with equal attribute values are defined as an equivalence class, which is also considered as an information granule [42]. In fuzzy contexts, a similarity relation can induce a fuzzy information granule [38] (also referred to as fuzzy granule or granule for simplicity), which is constructed by aggregating similar fuzzy points.

Definition 4. Let (U, A) be an information system, $\forall B \subseteq A$, the fuzzy granular structure of U induced by $\mathcal{R}_B$ is defined as

$$G_U(\mathcal{R}_B) = \{[x_1]_B, [x_2]_B, \dots, [x_n]_B\}, \tag{2}$$

where $[x_i]_B = \left(r_{i1}^B, r_{i2}^B, \dots, r_{in}^B\right)$ is employed as the fuzzy granule with x_i as its center.

Obviously, The fuzzy granule $[x_i]_B$ is a special fuzzy set on U, and $[x_i]_B(x_j) = \mathcal{R}_B(x_i, x_j) = r_{ij}^B$. If $\mathcal{R}_B(x_i, x_j) = 1$, then it suggests that x_j certainly belongs to $[x_i]_B$; If $\mathcal{R}_B(x_i, x_j) = 0$, then x_j definitely does not belong to $[x_i]_B$. The cardinality of the fuzzy rough granule $[x_i]_B$ is calculated by

$$\left|[x_i]_B\right| = \sum_{j=1}^{n} \mathcal{R}_B(x_i, x_j). \tag{3}$$

We can easily obtain $1 \le \left|[x_i]_B\right| \le n$.

Example 1. Given an information system (U, A), $U = \{x_1, x_2, x_3\}$, $B = \{a_1, a_2\} \subseteq A$ is a subset of attributes from A, the fuzzy similarity matrices on U induced from the attribute a_1 and a_2 respectively are

$$M(\mathcal{R}_{a_1}) = \begin{pmatrix} 1 & 0 & 0.8 \\ 0 & 1 & 0 \\ 0.8 & 0 & 1 \end{pmatrix}, \; M(\mathcal{R}_{a_2}) = \begin{pmatrix} 1 & 0.8 & 0.7 \\ 0.8 & 1 & 0 \\ 0.7 & 0 & 1 \end{pmatrix}.$$

The fuzzy relation matrix w.r.t. B by conjunction operation is

$$M(\mathcal{R}_B) = \begin{pmatrix} 1 & 0 & 0.7 \\ 0 & 1 & 0 \\ 0.7 & 0 & 1 \end{pmatrix}.$$

The fuzzy granule $[x_1]_B$ generated by $\mathcal{R}_B$ centered at x_1 is $(1, 0, 0.7)$, and $\left|[x_1]_B\right| = 1.7$. Similarly, $[x_2]_B = (0, 1, 0)$, $\left|[x_2]_B\right| = 1$, $[x_3]_B = (0.7, 0, 1)$, $\left|[x_3]_B\right| = 1.7$.

Fuzzy approximation [5] is an important concept of FRS, which provides a way to represent the ambiguous boundary regions between fuzzy sets. In these regions, the exact membership of elements is uncertain but can be approximated based on the available information. By leveraging fuzzy approximations, one can handle uncertain or imprecise data effectively. The definitions of upper and lower fuzzy approximations are as follows.

Definition 5. Let $\mathcal{R}_B$ be a fuzzy similarity relation on U. $\forall x \in U$ and $\mathcal{X} \in \mathcal{F}(U)$, the lower approximation $\underline{\mathcal{R}_B}\mathcal{X}$ and upper approximation $\overline{\mathcal{R}_B}\mathcal{X}$ of the fuzzy set $\mathcal{X}$ are a pair of fuzzy sets on U whose membership functions respectively are

$$\underline{\mathcal{R}_B}\mathcal{X}(x) = \inf_{y \in U} \max\left\{1 - \mathcal{R}_B(x, y), \mathcal{X}(y)\right\}, \tag{4}$$

$$\overline{\mathcal{R}_B}\mathcal{X}(x) = \sup_{y \in U} \min\left\{\mathcal{R}_B(x, y), \mathcal{X}(y)\right\}. \tag{5}$$

Example 2. Following Example 1, let $\mathcal{X} = (0.2, 0.5, 0.8)$ be a fuzzy set on U, then the fuzzy lower approximation of $\mathcal{X}$ with regard to $\mathcal{R}_B$ can be computed using Eq. (4) as

$\max\left\{1 - \mathcal{R}_B(x_1, y), \mathcal{X}(y)\right\}\Big|_{y=x_1} = \max\{0, 0.2\} = 0.2,$

$\max\left\{1 - \mathcal{R}_B(x_1, y), \mathcal{X}(y)\right\}\Big|_{y=x_2} = \max\{1, 0.5\} = 1,$

$\max\left\{1 - \mathcal{R}_B(x_1, y), \mathcal{X}(y)\right\}\Big|_{y=x_3} = \max\{0.3, 0.8\} = 0.8,$

then we have $\underline{\mathcal{R}_B}\mathcal{X}(x_1) = \inf_{y \in U} \max\left\{1 - \mathcal{R}_B(x_1, y), \mathcal{X}(y)\right\} = 0.2$.

Similarly, we can obtain $\underline{\mathcal{R}_B}\mathcal{X}(x_2) = 0.5$, $\underline{\mathcal{R}_B}\mathcal{X}(x_3) = 0.3$.

The fuzzy upper approximation of $\mathcal{X}$ with regard to $\mathcal{R}_B$ can be computed using Eq. (5) as

$\min\left\{\mathcal{R}_B(x_1, y), \mathcal{X}(y)\right\}\Big|_{y=x_1} = \min\{1, 0.2\} = 0.2,$

$\min\left\{\mathcal{R}_B(x_1, y), \mathcal{X}(y)\right\}\Big|_{y=x_2} = \min\{0, 0.5\} = 0,$

$\min\left\{\mathcal{R}_B(x_1, y), \mathcal{X}(y)\right\}\Big|_{y=x_3} = \min\{0.7, 0.8\} = 0.7,$

then we have $\overline{\mathcal{R}_B}\mathcal{X}(x_1) = \sup_{y \in U} \min\left\{\mathcal{R}_B(x_1, y), \mathcal{X}(y)\right\} = 0.7$.

Similarly, we can obtain $\overline{\mathcal{R}_B}\mathcal{X}(x_2) = 0.5$, $\overline{\mathcal{R}_B}\mathcal{X}(x_3) = 0.8$.

3.2. Multi-scale information system

An information system that takes various values for some attributes at multiple scale levels is referred to as a multi-scale information system (MIS). The notion of MIS is first introduced by Wu and Leung [13], which is defined as follows.

Definition 6. Let (U, A) be an information system, if A has a set of scales $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_s\}(s > 1)$, then (U, A^Λ) is called a multi-scale information system (MIS), which is defined as

$$\text{MIS} = \left(U, A^\Lambda\right) = \left(U, a_k^\lambda \middle| k = 1, 2, \dots, m; \lambda \in \Lambda\right). \tag{6}$$

4. Methodology

In this section, we first present a method for information fusion based on multi-scale fuzzy granules (MFG) which integrates the fuzzy information from multiple scale levels. Then an outlier detection algorithm based on fuzzy approximation is presented. Finally, we provide a concrete example to illustrate how the proposed method is working. Fig. 1 depicts the overall framework of our method MFIOD.

4.1. Information fusion based on multi-scale fuzzy granules

In an MIS, objects may have various values across multiple scale levels. If multi-scale information is fused properly, the performance and reliability of the detection algorithm will have the potential to improve. In the following, we begin with the definition of multi-scale fuzzy relation, and then construct multi-scale fuzzy granules (referred to as fuzzy granules for simplicity, unless confusion arises) by integrating single-scale fuzzy granules from each scale level.

During data preprocessing, we perform min–max standardization on each numerical attribute to transform their values into the range of 0 and 1, while symbolic attributes remain unaltered. Subsequently, we apply a distance-based measure as described by Wang et al. (2019) [43] to construct the single-scale fuzzy relation, which serves to represent objects with diverse types of values.

Given a multi-scale information system (U, A^Λ) with a scale set Λ, let f_i^k and f_j^k be the value of attribute a_k for object x_i and x_j, respectively. The distance between x_i and x_j on the attribute a_k is calculated as

$$d_{ij}^k = \begin{cases} 0, & \text{if } f_i^k = f_j^k, a_k \text{ is nominal;} \\ 1, & \text{if } f_i^k \neq f_j^k, a_k \text{ is nominal;} \\ |f_i^k - f_j^k|, & \text{if } a_k \text{ is numerical.} \end{cases} \tag{7}$$

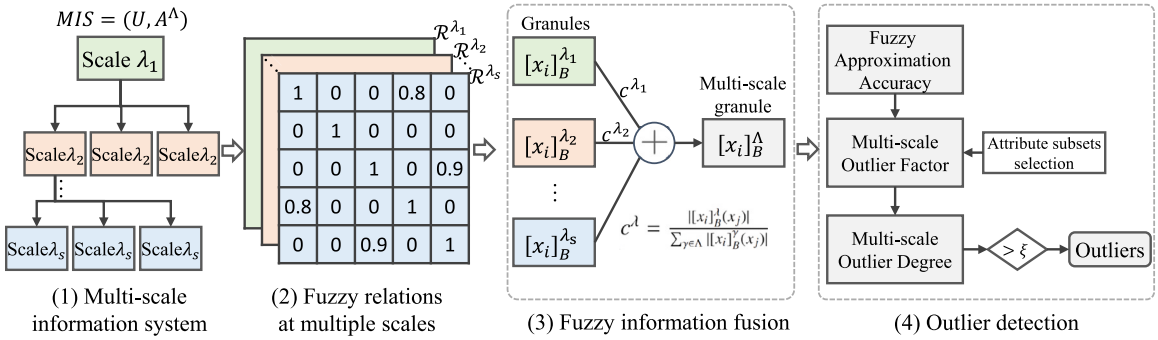

Fig. 1. Overall framework of MFIOD.

Then their distance on the attribute subset B is computed as $d_{ij}^{B} = \frac{1}{\sqrt{|B|}}\sqrt{\sum_{k\in B}(d_{ij}^{k})^2}$. For each scale $\lambda \in \Lambda$, the single-scale fuzzy similarity relation between x_i and x_j is computed as

$$\mathcal{R}_B^{\lambda}(x_i, x_j) = \begin{cases} 1 - d_{ij}^{B}, & d_{ij}^{B} \le \lambda; \\ 0, & d_{ij}^{B} > \lambda. \end{cases} \tag{8}$$

Since $\mathcal{R}_B^{\lambda}(x_i, x_j)$ exhibits both reflexivity: $\mathcal{R}_B^{\lambda}(x_i, x_i) = 1$ and symmetry: $\mathcal{R}_B^{\lambda}(x_i, x_j) = \mathcal{R}_B^{\lambda}(x_j, x_i)$, it follows that $\mathcal{R}_B^{\lambda}(x_i, x_j)$ qualifies as a fuzzy similarity relation.

Definition 7. Let (U, A^{Λ}) be an MIS, $B \subseteq A$, the multi-scale fuzzy relation $\mathcal{R}_B^{\Lambda}$ is induced by B at the scale set Λ, and its membership function is defined as

$$\mathcal{R}_B^{\Lambda}(x_i, x_j) = \sum_{\lambda\in\Lambda} c^{\lambda}\mathcal{R}_B^{\lambda}(x_i, x_j), \tag{9}$$

where $c^{\lambda} = \frac{\sum_{y\in U}\mathcal{R}_B^{\lambda}(x_i, y)}{\sum_{\gamma\in\Lambda}\sum_{y\in U}\mathcal{R}_B^{\gamma}(x_i, y)}$ is the weight coefficient of each scale level.

Proposition 1. *For $\Lambda = \{\lambda_1, \lambda_2\}$, if $\lambda_1 \le \lambda_2$, then $\mathcal{R}_B^{\lambda_1} \subseteq \mathcal{R}_B^{\Lambda} \subseteq \mathcal{R}_B^{\lambda_2}$.*

Proof. For any attribute subset B, $\forall x_i, x_j \in U$, if $\lambda_1 \le \lambda_2$, we have $\mathcal{R}_B^{\lambda_1}(x_i, x_j) \le \mathcal{R}_B^{\lambda_2}(x_i, x_j)$ by Eq. (8). Given $\Lambda = \{\lambda_1, \lambda_2\}$, we can obtain $c^{\lambda_1} + c^{\lambda_2} = 1$, and $\mathcal{R}_B^{\Lambda}(x_i, x_j) = c^{\lambda_1}\mathcal{R}_B^{\lambda_1}(x_i, x_j) + c^{\lambda_2}\mathcal{R}_B^{\lambda_2}(x_i, x_j)$ by Definition 7. Therefore, $\mathcal{R}_B^{\lambda_1}(x_i, x_j) = \frac{1}{c^{\lambda_1}}\mathcal{R}_B^{\Lambda}(x_i, x_j) - \frac{c^{\lambda_2}}{c^{\lambda_1}}\mathcal{R}_B^{\lambda_2}(x_i, x_j) \le \frac{1}{c^{\lambda_1}}\mathcal{R}_B^{\Lambda}(x_i, x_j) - \frac{c^{\lambda_2}}{c^{\lambda_1}}\mathcal{R}_B^{\lambda_1}(x_i, x_j)$, then $(1 + \frac{c^{\lambda_2}}{c^{\lambda_1}})\mathcal{R}_B^{\lambda_1}(x_i, x_j) \le \frac{1}{c^{\lambda_1}}\mathcal{R}_B^{\Lambda}(x_i, x_j)$. Hence, $\mathcal{R}_B^{\lambda_1} \subseteq \mathcal{R}_B^{\Lambda}$. Similarly, we have $\mathcal{R}_B^{\Lambda} \subseteq \mathcal{R}_B^{\lambda_2}$. □

A multi-scale fuzzy relation derives a collection of multi-scale fuzzy granules, which are constructed by drawing together similar objects on a subset of attributes at multiple scale levels.

Definition 8. Let (U, A^{Λ}) be an MIS with a scale set Λ, $\forall B \subseteq A$, the multi-scale fuzzy granule $[x_i]_B^{\Lambda}$ centered at x_i can be generated by the multi-scale fuzzy relation $\mathcal{R}_B^{\Lambda}$, and its membership function is defined as

$$[x_i]_B^{\Lambda}(x_j) = \sum_{\lambda\in\Lambda} c^{\lambda}[x_i]_B^{\lambda}(x_j), \tag{10}$$

where $c^{\lambda} = \frac{|[x_i]_B^{\lambda}(x_j)|}{\sum_{\gamma\in\Lambda}|[x_i]_B^{\gamma}(x_j)|}$ is the weight coefficient of each scale level.

Proposition 2. *For $\Lambda = \{\lambda_1, \lambda_2\}$, if $\lambda_1 \le \lambda_2$, then $[x_i]_B^{\lambda_1} \subseteq [x_i]_B^{\Lambda} \subseteq [x_i]_B^{\lambda_2}$.*

Proof. Proposition 2 is easily derived through a process similar to that of Proposition 1. □

Proposition 2 shows the fuzzy granules increase monotonically with the scale size, and the combination of multiple scales creates a new fuzzy granule that integrates information from multiple scale levels. Each multi-scale fuzzy granule indicates a pile of objects that are drawn together by their similarities measured at multiple scale levels. The high cardinality of a granule $[x_i]_B^{\Lambda}$ implies that x_i is frequently associated with other objects, therefore, x_i is more likely to be a majority.

4.2. Outlier detection based on fuzzy approximation

Outlier detection algorithms usually specify an outlierness value (also called outlier degree) for each object, which can be used to compare with other objects. If the outlier degree of an object is relatively high, it will be more likely to be an outlier. In this subsection, we begin with the definition of fuzzy approximation accuracy, then introduce two concepts for the quantification of the outlierness of an object, that is, the multi-scale outlier factor (MOF) and the multi-scale outlier degree (MOD). MOF is taken to measure the outlier degree of the multi-scale fuzzy granule (MFG), while MOD indicates the outlier degree of each object in data.

4.2.1. Fuzzy approximation accuracy

Let $[x_i]_B^{\Lambda}$ be a MFG with x_i as its center, $\forall P \subseteq A$, the lower approximation $\underline{\mathcal{R}_P}[x_i]_B^{\Lambda}$ and upper approximation $\overline{\mathcal{R}_P}[x_i]_B^{\Lambda}$ of $[x_i]_B^{\Lambda}$ w.r.t. $\mathcal{R}_P$ can be generated, and their membership functions are respectively as follows.

$$\begin{aligned} \underline{\mathcal{R}_P}[x_i]_B^{\Lambda}(x_j) &= \inf_{y\in U} \max\left\{1 - \mathcal{R}_P(x_j, y), [x_i]_B^{\Lambda}(y)\right\}, \\ \overline{\mathcal{R}_P}[x_i]_B^{\Lambda}(x_j) &= \sup_{y\in U} \min\left\{\mathcal{R}_P(x_j, y), [x_i]_B^{\Lambda}(y)\right\}. \end{aligned} \tag{11}$$

The two approximations of $[x_i]_B$ are used to express the degree of x_j certainly belonging to $[x_i]_B$ and the degree of x_j possibly belonging to $[x_i]_B$, respectively. They are further employed to define the fuzzy approximation accuracy that expresses the "quality" of the approximations.

Definition 9. Given a multi-scale fuzzy granule $[x_i]_B^{\Lambda}$ centered at x_i, $\forall P \subseteq A$, the approximation accuracy of $[x_i]_B^{\Lambda}$ w.r.t. P is defined as

$$\alpha_P([x_i]_B^{\Lambda}) = \frac{\left|\underline{\mathcal{R}_P}[x_i]_B^{\Lambda}\right|}{\left|\overline{\mathcal{R}_P}[x_i]_B^{\Lambda}\right|}. \tag{12}$$

Obviously, $0 \le \alpha_P([x_i]_B^{\Lambda}) \le 1$. The fuzzy approximation accuracy quantifies how difficult a fuzzy granule can be approximated by other granules or attributes. If $\alpha_P([x_i]_B^{\Lambda})$ is low, the granule $[x_i]_B^{\Lambda}$ is hard to be approximated by P.

4.2.2. Multi-scale outlier factor

The core thinking of MOF is that if a fuzzy granule exhibits uniquely, it would be difficult to be approximated by many other granules. Therefore, we can use MOF to measure the outlier degree of a fuzzy granule, which is defined as

Definition 10. Let (U, A^{Λ}) be an MIS, $B, P \subseteq A$, $\forall x_i \in U$, the multi-scale outlier factor (MOF) of the multi-scale fuzzy granule $[x_i]_B^{\Lambda}$ is calculated by

$$MOF\left([x_i]_B^{\Lambda}\right) = 1 - \frac{1}{|U|}\left|[x_i]_B^{\Lambda}\right| \cdot \alpha_P([x_i]_B^{\Lambda}). \tag{13}$$

In the above definition, all attribute subsets of A can be used for constructing granules, and it would be impractical or infeasible for inputs of even moderate size. Recent research addresses this problem by using some heuristic means. For example, Jiang et al. [34] create two sequences of attribute subsets based on their information entropy. Yuan et al. [38] simply ignore subsets with the number of elements greater than two. This paper introduces a new approach for determining the attribute subset B based on the standard deviation of the values of each attribute. Specifically, let δ_k be the standard deviation of attribute a_k, we first sort all attributes in ascending order of their deviations, and then divide them into M equal-width bins. The range of the lth bin (i.e., the lth attribute subset B_l) is denoted as

$$Interval_{B_l} = \left[\delta_{\min} + (l-1) \times BinWidth, \delta_{\min} + l \times BinWidth\right), \tag{14}$$

where $BinWidth = (\delta_{\max} - \delta_{\min})/M$ is the width of each bin, $\delta_{\min}$ and $\delta_{\max}$ are the minimum value and maximum value of standard deviations of all attributes, respectively. In this paper, the default number of attribute bins is set to 20 (i.e., $M = 20$). Let $\bar{D}_{B_l} = \frac{1}{|U|^2}\sum_{i,j} d_{ij}^{B_l}$ be the average distance associated with an attribute bin B_l, we choose the three smallest average distances as the default attribute scales.

4.2.3. Multi-scale outlier degree

As outliers generally exhibit more unique than normal data, the MFG of outliers would be harder to approximate by the others, and the average MOF of outliers should be relatively high. On this base, the MOD of each instance is obtained by integrating the MOF of a set of granules in the form of weighted summation.

Definition 11. Let $\left\{B_l | l = 1, 2, \ldots, M; B_l \subseteq A\right\}$ be the set of attribute subset (bin), $\forall x_i \in U$, the multi-scale outlier degree (MOD) of x_i is defined as

$$MOD(x_i) = \frac{1}{M}\sum_{l=1}^{M}\left(MOF\left([x_i]_{B_l}^{\Lambda}\right) \cdot W_{B_l}(x_i)\right), \tag{15}$$

where $W_{B_l} : U \to [0, 1]$ is a weight coefficient computed by $W_{B_l}(x_i) = 1 - \sqrt[3]{\left|[x_i]_{B_l}^{\Lambda}\right|/|U|}$.

Given an object x_i, the MOF of each MFG of x_i is computed, and each MOF is weighted by a coefficient $W_{B_l}(x_i)$ with the assumption that outliers are usually a small number of objects in data. In other words, the minority objects should have higher weights than the majority of data. Specifically, $\forall x_i \in U$, if the cardinality of $[x_i]_{B_l}^{\Lambda}$ is relatively small, x_i is treated as a minority in U. Hence, a larger value of $MOD(x_i)$ implies a higher likelihood of x_i being an outlier.

4.2.4. Outlier detection algorithm

Practically, a threshold ξ is employed to filter out normal objects. This threshold is determined in such a way that it selects a specific proportion of objects in U as outliers. Then, we can obtain a binary output that determines the outliers in data.

Definition 12. Let ξ be a real-valued threshold, $\forall x_i \in U$, if $MOD(x_i) > \xi$, then x_i is regarded as an outlier.

Algorithm 1 elaborates the whole procedure of the proposed outlier detection method. MFIOD first groups all attributes into several bins (Lines 1 & 2) then calculates multi-scale fuzzy relation matrices of each attribute bin (Lines 4 to 8), and then constructs multi-scale fuzzy granules of each object w.r.t. each attribute bin (Lines 10 to 12). Later, the outlier factor of each fuzzy granule is computed (Lines 13 & 14).

Algorithm 1: MFIOD

Input: A multi-scale information system (U, A^{Λ}), the threshold ξ.
Output: Outlier set O

1 $O \leftarrow \emptyset$;
2 Calculate the standard deviation for each attribute;
3 Group all attributes into M bins using Eq. (14);
4 **for** $l \leftarrow 1$ *to* M **do**
5 **for** *each* λ *in* Λ **do**
6 Compute fuzzy relation matrix $M(\mathcal{R}_{B_l}^{\lambda})$ using Eq. (8);
7 **end**
8 Compute multi-scale relation matrix $M(\mathcal{R}_{B_l}^{\Lambda})$ using Eq. (9);
9 **end**
10 **for** *each* $x_i \in U$ **do**
11 **for** $l \leftarrow 1$ *to* M **do**
12 Compute multi-scale fuzzy granules $[x_i]_{B_l}^{\Lambda}$ using Eq. (10);
13 Compute fuzzy approximation accuracy $\alpha_P([x_i]_{B_l}^{\Lambda})$ using Eqs. (11) and (12);
14 Compute outlier factor $MOF([x_i]_{B_l}^{\Lambda})$ using Eq. (13);
15 **end**
16 Compute multi-scale outlier degree $MOD(x_i)$ using Eq. (15);
17 **if** $MOD(x_i) > \xi$ **then**
18 $O = O \cup \{x_i\}$;
19 **end**
20 **end**
21 **return** O.

Table 3
An example of an information system.

U	a_1	a_2	a_3	a_1	a_2	a_3
x_1	0.7	9	A	1	1	0
x_2	0.3	6	B	0.2	0.571	1
x_3	0.5	2	A	0.6	0	0
x_4	0.2	3	B	0	0.143	1
x_5	0.4	7	A	0.4	0.714	0
x_6	0.6	3	B	0.8	0.143	1

Finally, MOD scores of all data points are obtained and outliers are determined accordingly (Lines 16 to 19).

The total loops of MFIOD is $n \times s + n^2$, $(n \gg s)$. Therefore, the worst time complexity of MFIOD is $O(n^2)$. MFIOD has a lower or similar computational complexity compared to the other four algorithms GrC ($O(mn^2)$), IE ($O(m^2 n \log n)$), ODGrCR ($O(m^2(mn + e))$), and FRGOD ($O(mn(m+n))$), which are also based on Granular computing. To illustrate the computation process of the proposed method more clearly, a toy example is given as follows.

Example 3. The left side of Table 3 represents an MIS, where $U = \{x_1, x_2, \ldots, x_6\}$ is the set of objects, $A = \{a_1, a_2, a_3\}$ is the attribute set with the scale set $\Lambda = \{\lambda_1, \lambda_2\}$ and $\lambda_1 = 0.2$, $\lambda_2 = 0.5$. We first apply min–max normalization to all values. The results are on the right side of Table 3. Then we compute the standard deviation for each attribute, $\delta_{a_1} \approx 0.342$, $\delta_{a_2} \approx 0.360$, $\delta_{a_3} = 0.5$.

Let the number of attribute bins be 2, then we obtain two bins: $B_1 = \{a_1, a_2\}$, $B_2 = \{a_3\}$.

Next, we construct the fuzzy relation matrices at two different scales, i.e., λ_1 and λ_2 for each attribute subset (attribute bin). By computing the normalized Euclidean distance of each pair of objects under two attribute bins, we obtain two distance matrices. Taking the

distance between x_1 and x_5 under attribute bin B_1 as example, we have

$$d_{15}^{B_1} = \frac{1}{\sqrt{|B_1|}}\sqrt{(f_{x_1}^{a_1} - f_{x_5}^{a_1})^2 + (f_{x_1}^{a_2} - f_{x_5}^{a_2})^2} = \frac{1}{\sqrt{2}}\sqrt{(1-0.4)^2 + (1-0.714)^2} \approx 0.470$$

By using Eq. (8), we can obtain the fuzzy relation between x_1 and x_5 induced by B_1. As $d_{15}^{B_1} > \lambda_1$ and $d_{15}^{B_1} < \lambda_2$, we have $\mathcal{R}_{B_1}^{\lambda_1}(x_1, x_5) = 0$ and $\mathcal{R}_{B_1}^{\lambda_2}(x_1, x_5) = 1 - d_{15}^{B_1} = 0.53$. Similarly, we have the fuzzy relation matrices at two different scales as follows

$$M(\mathcal{R}_{B_1}^{\lambda_1}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0.826 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0.826 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0.826 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.826 & 0 & 0 & 1 \end{pmatrix},$$

$$M(\mathcal{R}_{B_1}^{\lambda_2}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0.53 & 0 \\ 0 & 1 & 0.507 & 0.666 & 0.826 & 0 \\ 0 & 0.507 & 1 & 0.564 & 0 & 0.826 \\ 0 & 0.666 & 0.564 & 1 & 0.507 & 0 \\ 0.53 & 0.826 & 0 & 0.507 & 1 & 0.507 \\ 0 & 0 & 0.826 & 0 & 0.507 & 1 \end{pmatrix},$$

$$M(\mathcal{R}_{B_2}^{\lambda_1}) = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix} = (\mathcal{R}_{B_2}^{\lambda_2}).$$

In order to fuse the fuzzy relations of the two scales, we calculate below multi-scale relation matrices induced by each attribute subset in the bins using Eq. (10):

$$M(\mathcal{R}_{B_1}^{\Lambda}) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0.321 & 0 \\ 0 & 1 & 0.315 & 0.414 & 0.826 & 0 \\ 0 & 0.311 & 1 & 0.346 & 0 & 0.826 \\ 0 & 0.487 & 0.413 & 1 & 0.371 & 0 \\ 0.344 & 0.826 & 0 & 0.329 & 1 & 0.329 \\ 0 & 0 & 0.826 & 0 & 0.284 & 1 \end{pmatrix},$$

$M(\mathcal{R}_{B_2}^{\Lambda}) = M(\mathcal{R}_{B_1}^{\lambda_1})$.

Next, we compute the fuzzy approximation accuracy. First, we calculate the fuzzy relation matrix $M(\mathcal{R}_P)$ induced by all attributes (i.e., $P = A$) as

$$M(\mathcal{R}_P) = \begin{pmatrix} 1 & 0.22 & 0.378 & 0.045 & 0.616 & 0.231 \\ 0.22 & 1 & 0.296 & 0.727 & 0.405 & 0.574 \\ 0.378 & 0.296 & 1 & 0.322 & 0.572 & 0.405 \\ 0.045 & 0.727 & 0.322 & 1 & 0.296 & 0.538 \\ 0.616 & 0.405 & 0.572 & 0.296 & 1 & 0.296 \\ 0.231 & 0.574 & 0.405 & 0.538 & 0.296 & 1 \end{pmatrix},$$

and use $M(\mathcal{R}_P)$ to approximate each attribute subset. Herein, x_1 is taken as an example. From the above multi-scale relation matrix $M(\mathcal{R}_{B_1}^{\Lambda})$, we can easily get $[x_1]_{B_1}^{\Lambda} = (1, 0, 0, 0, 0.321, 0)$, then the fuzzy approximation can be computed by using Eq. (11) (See Example 2) as $\underline{\mathcal{R}_P}[x_1]_{B_1}^{\Lambda} = (0.384, 0, 0, 0, 0.321, 0)$, $\overline{\mathcal{R}_P}[x_1]_{B_1}^{\Lambda} = (1, 0.321, 0.378, 0.296, 0.616, 0.296)$.

Then we can get the approximation accuracy of the fuzzy granule with x_1 as its center with regard to $\mathcal{R}_P$ as

$$\alpha_P([x_1]_{B_1}^{\Lambda}) = \frac{\left|\underline{\mathcal{R}_P}[x_1]_{B_1}^{\Lambda}\right|}{\left|\overline{\mathcal{R}_P}[x_1]_{B_1}^{\Lambda}\right|} \approx \frac{0.705}{2.907} \approx 0.242.$$

By Eq. (13), the outlier factor of $[x_1]_{B_1}^{\Lambda}$ and its related weight are calculated as

$$MOF\left([x_1]_{B_1}^{\Lambda}\right) = 1 - \frac{1}{|U|} \times \left|[x_1]_{B_1}^{\Lambda}\right| \times \alpha_P([x_1]_{B_1}^{\Lambda}) \approx 1 - \frac{1}{6} \times 1.321 \times 0.242 \approx 0.947,\ W_{B_1}(x_1) = 1 - \sqrt[3]{\left|[x_1]_{B_1}^{\Lambda}\right| / |U|} \approx 1 - \sqrt[3]{1.321/6} \approx 0.396.$$

Similarly, we have $MOF\left([x_1]_{B_2}^{\Lambda}\right) \approx 0.763, W_{B_2}(x_1) \approx 0.206$. Then the multi-scale outlier degree of each object is calculated using Eq. (15) as

$$MOD(x_1) = \frac{1}{2}\sum_{k=1}^{2}\left(MOF\left([x_i]_{B_l}^{\Lambda}\right) \cdot W_{B_l}(x_i)\right) \approx \frac{1}{2} \times (0.947 \times 0.396 + 0.763 \times 0.206) \approx 0.266,\ MOD(x_2) \approx 0.185,\ MOD(x_3) \approx 0.186,\ MOD(x_4) \approx 0.192,\ MOD(x_5) \approx 0.169,\ MOD(x_6) \approx 0.211.$$

After comparing the MOD sores of all objects, we can see that the outlier factor of x_1 is obviously higher than others, which implies that x_1 is most likely to be an outlier.

5. Experiments

This section systematically conducts experiments to evaluate our proposed algorithm on 15 public datasets,[2] then analyzes the parameter sensitivity. All the data sets and source codes are publicly available online.[3]

5.1. Datasets

The statistics of the experimental datasets are listed in Table 4. Following previous works [38,44], this paper also adopts the down-sampling method which randomly removes some objects in a particular class to produce an outlier set. This is to simulate the scenario of outlier detection in real-world applications, where outliers are rare and difficult to find. Next, the maximum probability method is taken to handle the missing values, i.e., to fill in the blanks with the most frequently occurring values on other samples. In addition, all attribute values are transformed into the interval of $[0, 1]$ by adopting the min–max normalization.

5.2. Comparison methods

We take the following 11 comparison methods [4] to evaluate the performance of our proposed method:

- GrC (2008) [33]: A GrC-based method, which uses overlap distance to calculate pairwise distance.
- IForest (2008) [21]: An ensemble model which isolates objects by randomly selecting a feature and a split value between the maximum and minimum of the selected feature.
- LoOP (2009) [25]: A hybrid model, which combines density-based local outlier scores with a probabilistic approach.
- LDOF (2009) [22]: A distance-based algorithm that exploits the relative location of a sample to its neighbors to measure the outlier degree.
- IE (2010) [34]: A rough sets-based algorithm, which adopts information entropy to measure the uncertainty of data.
- WDOD (2014) [26]: A weighted density based algorithm for categorical attributes.
- ODGrCR (2015) [35]: A rough sets-based model, which extends to apply rough approximation accuracy in outlier detection.
- ECOD (2022) [20]: A statistic-based model that computes empirical cumulative distribution for each dimension of data in a non-parametric fashion.
- ROD (2022) [45]: A subspace learning-based method that decomposes the attribute space into combinations of subspaces for computing outlier score.

[2] https://github.com/BELLoney/Outlier-detection
[3] https://github.com/ChenBaiyang/MFIOD
[4] https://github.com/yzhao062/pyod

Table 4
The statistics of the datasets used in the experiments.

No.	Dataset	Abbr.	#Objects	#Attributes	#Outliers	%Outlier	Type
1	Annthyroid	Ann	7200	6	534	7.4%	Numerical
2	Cardiotocography_2and3_33_variant1	Card	1688	21	33	2.0%	Numerical
3	Diabetes_tested_positive_26_variant1	Diab	526	8	26	4.9%	Numerical
4	Ecoli	Ecoli	336	7	9	2.7%	Numerical
5	German_1_14_variant1	Germ	714	20	14	2.0%	Hybrid
6	Heart270_2_16_variant1	Heart	166	13	16	9.6%	Hybrid
7	Hepatitis_2_9_variant1	Hepa	94	19	9	9.6%	Hybrid
8	Ionosphere_b_24_variant1	Iono	249	34	24	9.6%	Numerical
9	Lymphography	Lymp	148	18	6	4.1%	Nominal
10	Pageblocks_1_258_variant1	Page	5171	10	258	5.0%	Numerical
11	Pima_TRUE_55_variant1	Pima	555	9	55	9.9%	Numerical
12	Sonar_M_10_variant1	Sonar	107	60	10	9.3%	Numerical
13	Wbc_malignant_39_variant1	Wbc	483	9	39	8.1%	Numerical
14	Wdbc_M_39_variant1	Wdbc	396	31	39	9.8%	Numerical
15	Yeast_ERL_5_variant1	Yeast	1141	8	5	0.4%	Numerical

- DCROD (2022) [29]: A density-based method, which combines kernel density estimation with extended neighbors to calculate the local density.
- FRGOD (2022) [38]: A fuzzy rough sets-based model that directly deals with mixed attribute data.

Among them, GrC, IE, WDOD and ODGrCR are only applicable to categorical values and require discretization operation to transform numerical data into discrete values. Following [38], we adopt Weka's discretization algorithm to convert all numbers into discrete values with three categories. As LDOF, LoOP and DCROD involve the parameter k, we search the optimal value of k varying from 1 to 60 with step size 1. The number of base estimators for IForest is set to 100. For FRGOD, we use the method reported in their paper to tune the parameter λ. Our proposed algorithm MFIOD requires adjustment of attribute scales. Initially, we adopt a single attribute scale i.e., $\Lambda = \{\lambda_1\}$, with its value ranging in the interval of [0.05, 1] with the stepsize of 0.1. Once an optimal scale value is identified, it is retained, and then the next scale is selected similarly. This process continues until the maximum number of scales is reached. We empirically set the maximum number of scales to 3.

5.3. Evaluation metrics

This study evaluates all the comparison methods in terms of precision scores with a parameter t [38], i.e., $P(t)$. All samples in a dataset are ordered according to their outlier scores generated by a detection method. Given an order indicator t, samples with an order indicator smaller than t are considered normal. A larger $P(t)$ denotes a better detection result. In our experiments, we fix $t = |OT|$. Let $OS(t)$ be the set of outliers detected under a given t, OT denotes true outliers in the dataset, $P(t)$ is computed by

$$P(t) = \frac{|OS(t) \cap OT|}{|OS(t)|}. \tag{16}$$

Meanwhile, we also use the Receiver Operating Characteristic (ROC) curve to measure the comparison results. Given a threshold parameter T, the sample x is classified as "positive" if the outlier score $s(x) > T$, and "negative" otherwise. The ROC curve plots parametrically the False Positive Rate (FPR) versus the True Positive Rate (TPR) with T as the varying parameter. They are calculated respectively by

$$\mathrm{FPR}(T) = \frac{|OS'(T) - OT|}{|U - OT|}, \tag{17}$$

$$\mathrm{TPR}(T) = \frac{|OS'(T) \cap OT|}{|OT|}, \tag{18}$$

where $OS'(T)$ represents the set of true anomalies w.r.t. the threshold T. If the ROC of a prediction algorithm is closer to the upper left of the ROC space, its detection performance is better.

5.4. Experimental results

The overall experimental results are listed in Table 5. The table shows the $P(t)$ scores of comparison detection methods on 15 datasets. The best score on each dataset is in bold. One can observe that MFIOD outperforms or ties with the other methods on 10 datasets, such as Diab, Ecoli, Germ, Heart, Hepa, Lymp, Pima, Sonar, Wbc, and Yeast. Moreover, MFIOD achieves the highest average $P(t)$ score across all datasets. In addition, MFIOD is not the best performer on five datasets including Ann, Card, Iono, Page, and Wdbc. However, it ranks second on Card, Iono, and Page, with only slight differences from ECOD, DCROD, and IForest. This suggests that MFIOD is more versatile and robust than other detectors.

In terms of ROC, we can see the ROC curves from Fig. 2, MFIOD is closer to the upper left of the ROC space on 10 datasets, which is consistent with the $P(t)$ metric. However, on datasets Ann, Card, and Page, MFIOD's curves are lower or cross with the best performer, i.e., FRGOD, ROD, and IForest. This suggests that MFIOD is less effective than these methods on these datasets. This may be due to the different outlier distributions in these datasets, where MFIOD's assumption does not hold.

5.5. Parameter analyses

As the paper mainly focuses on discussing the potential of multi-scale data representation, the number of scales and the size of each scale are the key parameters for the MFIOD algorithm. We evaluate the performances of MFIOD w.r.t. one ($\Lambda = \{\lambda_1\}$), two ($\Lambda = \{\lambda_1, \lambda_2\}$) and three scales($\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$), respectively. The parameter λ_i controls the size of granules of data points. $P(t)$ is taken as the metric for analyses in this part. In this part, we try to explore below two questions:

(1) What is the optimal number of scales for the proposed method?

As shown in Table 6, when $|\Lambda| = 3$, MFIOD achieves the best scores on four datasets including Ecoli, Heart, Page, and Pima. Specifically, three-scale MFIOD beats that of one-scale by 11.11%, 6.25%, 2.33% and 1.82% on Ecoli, Heart, Page, and Pima, respectively. This demonstrates the effectiveness of our proposed multi-scale information fusion algorithm. When $|\Lambda| = 2$, MFIOD performs equally with $|\Lambda| = 3$ on two datasets, i.e. Card and Yeast. Concretely, two-scale MFIOD enjoys a 3.03% higher $P(t)$ score on Card than single-scale MFIOD, and 20% higher on Yeast, which also strongly explains the superiority of the multi-scale approach for outlier detection. When $|\Lambda| = 1$, MFIOD achieves equal scores with that of two-scales or three-scales on 8 datasets, i.e., Diab, Germ, Hepa, Iono, Lymp, Sonar, Wbc, and Wdbc. Also, MFIOD performs the best on only one dataset, i.e., Ann. This indicates these datasets are not sensitive to different scale settings.

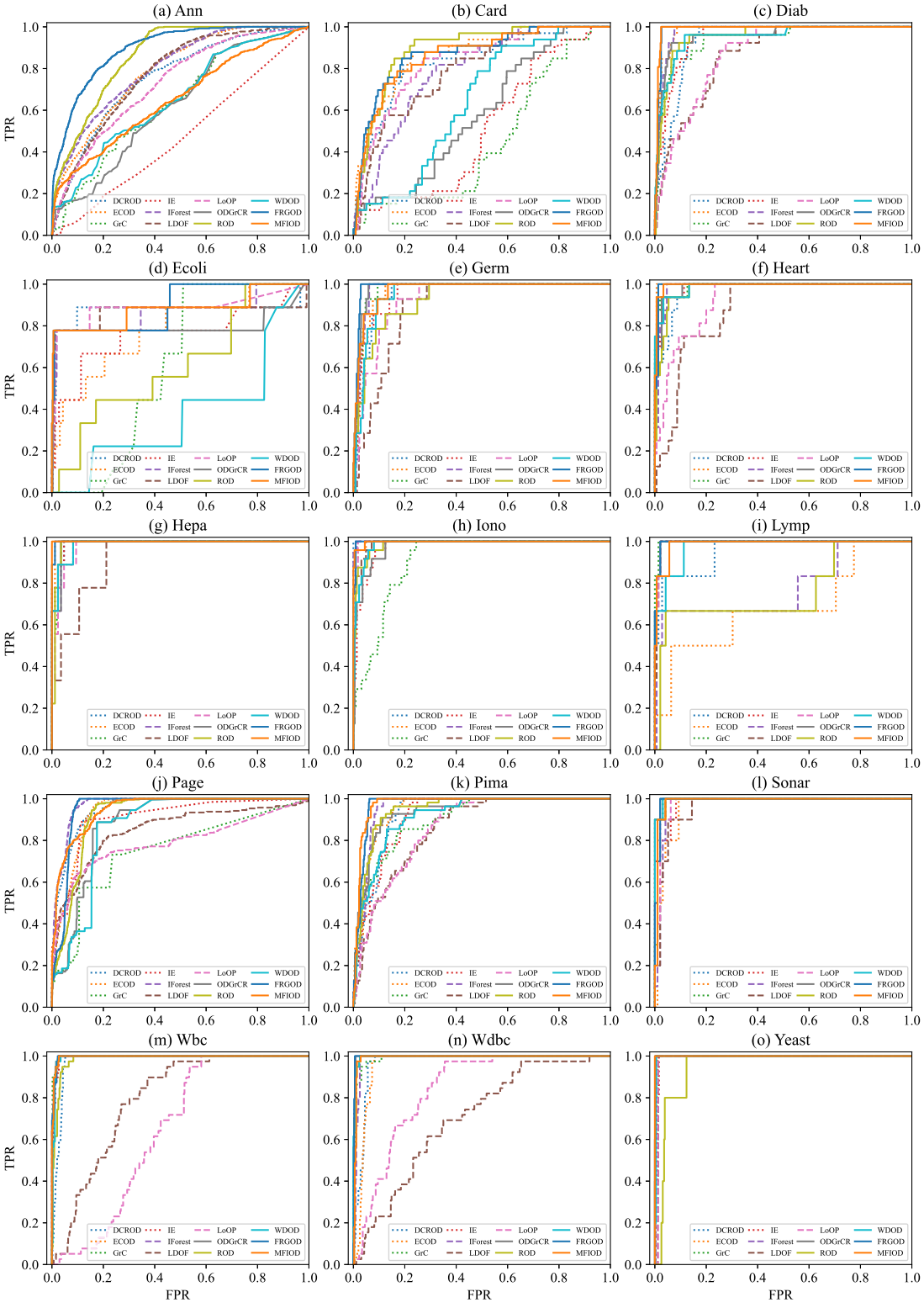

Fig. 2. ROC of all comparison methods on 15 experimental datasets.

Table 5
$P(t)$ (%) of comparison detection methods on the experimental datasets ($t = |OT|$).

Dataset	LDOF	ECOD	ROD	IForest	LoOP	WDOD	DCROD	GrC	IE	ODGrCR	FRGOD	MFIOD
Ann	23.22	30.71	35.21	33.15	23.60	16.48	22.66	15.36	5.62	16.48	**43.82**	26.59
Card	3.03	**27.27**	12.12	12.12	6.06	12.12	12.12	3.03	6.06	12.12	18.18	18.18
Diab	30.77	61.54	53.85	57.69	26.92	57.69	34.62	53.85	42.31	57.69	69.23	**76.92**
Ecoli	55.56	22.22	0.00	55.56	55.56	0.00	44.44	0.00	33.33	77.78	66.67	**77.78**
Germ	7.14	21.43	28.57	35.71	14.29	28.57	21.43	35.71	28.57	35.71	42.86	**42.86**
Heart	31.25	**93.75**	68.75	81.25	56.25	81.25	68.75	87.50	75.00	75.00	87.50	**93.75**
Hepa	55.56	88.89	77.78	66.67	77.78	77.78	77.78	77.78	77.78	66.67	88.89	**100.00**
Iono	83.33	91.67	83.33	95.83	87.50	79.17	**100.00**	41.67	70.83	70.83	91.67	95.83
Lymp	66.67	16.67	50.00	50.00	66.67	66.67	66.67	83.33	83.33	83.33	83.33	**83.33**
Page	43.02	39.92	27.52	**58.91**	41.09	18.60	53.88	17.83	43.41	17.05	29.84	58.53
Pima	41.82	56.36	60.00	54.55	40.00	50.91	49.09	54.55	45.45	56.36	65.45	**76.36**
Sonar	70.00	70.00	**90.00**	70.00	80.00	**90.00**	80.00	**90.00**	80.00	**90.00**	80.00	**90.00**
Wbc	17.95	89.74	74.36	84.62	5.13	87.18	61.54	89.74	84.62	89.74	87.18	**89.74**
Wdbc	20.51	61.54	92.31	79.49	30.77	92.31	61.54	87.18	**94.87**	**94.87**	**94.87**	89.74
Yeast	0.00	40.00	0.00	40.00	0.00	20.00	20.00	40.00	40.00	60.00	80.00	**100.00**
Average	36.66	54.11	50.25	58.37	40.77	51.92	51.63	51.83	54.08	60.24	68.63	**74.35**

Table 6
$P(t)$ of MFIOD w.r.t. multiple scale numbers ($t = |OT|$).

Dataset	$\|\Lambda\| = 1$	$\|\Lambda\| = 2$	$\|\Lambda\| = 3$
Ann	**26.59**	23.97	22.66
Card	15.15	**18.18**	**18.18**
Diab	**76.92**	**76.92**	**76.92**
Ecoli	66.67	66.67	**77.78**
Germ	**42.86**	**42.86**	**42.86**
Heart	87.50	87.50	**93.75**
Hepa	**100.00**	**100.00**	**100.00**
Iono	**95.83**	**95.83**	**95.83**
Lymp	**83.33**	**83.33**	**83.33**
Page	56.20	57.75	**58.53**
Pima	74.55	74.55	**76.36**
Sonar	**90.00**	**90.00**	**90.00**
Wbc	**89.74**	**89.74**	**89.74**
Wdbc	**89.74**	**89.74**	**89.74**
Yeast	80.00	**100.00**	**100.00**
Average	71.67	73.14	**74.38**

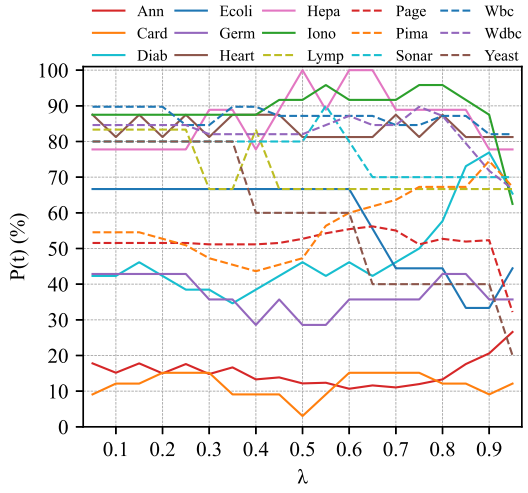

Fig. 3. $P(t)$ of MFIOD w.r.t. single-scale fuzzy granule ($t = |OT|$).

(2) What are the best scale sizes on the experimental datasets? To answer this question, we compare the $P(t)$ score w.r.t. various scale sizes of multi-scale settings. Fig. 3 depicts the results of MFIOD on 15 datasets w.r.t. a single-scale fuzzy granule. The parameter λ controls the scale size of fuzzy granules. A smaller scale size implies a more fine-grained analysis of the data. We can observe that the $P(t)$ first increases and then decreases on Diab, Hepa, Iono, Sonar, and Wdbc; while the datasets of Ann, Card, Germ, and Pima exhibit the opposite trend; for datasets Ecoli, Lymp, Page, and Yeast the $P(t)$ decrease monotonically with scale size λ. Moreover, the $P(t)$ of MFIOD changes little with λ on datasets Heart and Wbc. The optimal scale size of fuzzy granules depends on the characteristics of the datasets. Different datasets may have different patterns of λ.

The results of MFIOD w.r.t. two-scale fuzzy granules are shown in Fig. 4. The color gradient from blue to yellow represents the change in detection performance from low to high. The figure presents optimal results in some continuing areas on most datasets. Specifically, MFIOD has a large area of yellow color on Ecoli, Germ, Heart, Hepa, Lymp, Wbc, and Wdbc; for the remaining datasets, MFIOD shows a peak value at some small intermediate area. This suggests that certain combinations of two-scale sizes work well on these datasets, resulting in consistently high detection performance. Moreover, MFIOD has several valley areas on datasets Diab and Yeast, where the model's performance drops.

In the following, we investigate MFIOD's performance in three-scale fuzzy granule settings. Taking the results on Pima as an example, the 3D surface plot in Fig. 5 shows $P(t)$ score as a function of three-scale parameters λ_1, λ_2, and λ_3. The x-axis and y-axis represent λ_1 and λ_2, respectively. The z-axis represents the $P(t)$ score. Each subfigure in the plot corresponds to a fixed value of λ_3 from 0.7 to 0.95. The area with a yellow color indicates higher performance. One can observe that MFIOD performs optimally when the scale sizes are relatively large on Pima (around 0.9). The combination of scale size parameters leading to optimal performance varies across different datasets.

6. Conclusion

Aiming at the task of unsupervised outlier detection, this paper presents a multi-scale information fusion approach based on fuzzy granules and constructs an effective outlier detection algorithm MFIOD with the fuzzy rough sets theory. Specifically, MFIOD employs fuzzy approximation to characterize outliers in data, and computes the multi-scale outlier degree of each object based on their fuzzy approximation accuracy. Experimental results on nine public datasets show that our proposed algorithm can effectively integrate multi-scale information in data and achieve optimal performance for detecting outliers. However, the space complexity of the proposed method is relatively high and, therefore needs to be optimized in future work. In addition, integrating multi-scale information through other means (e.g., η-fuzzy similarity relation [46]) is also a potential direction.

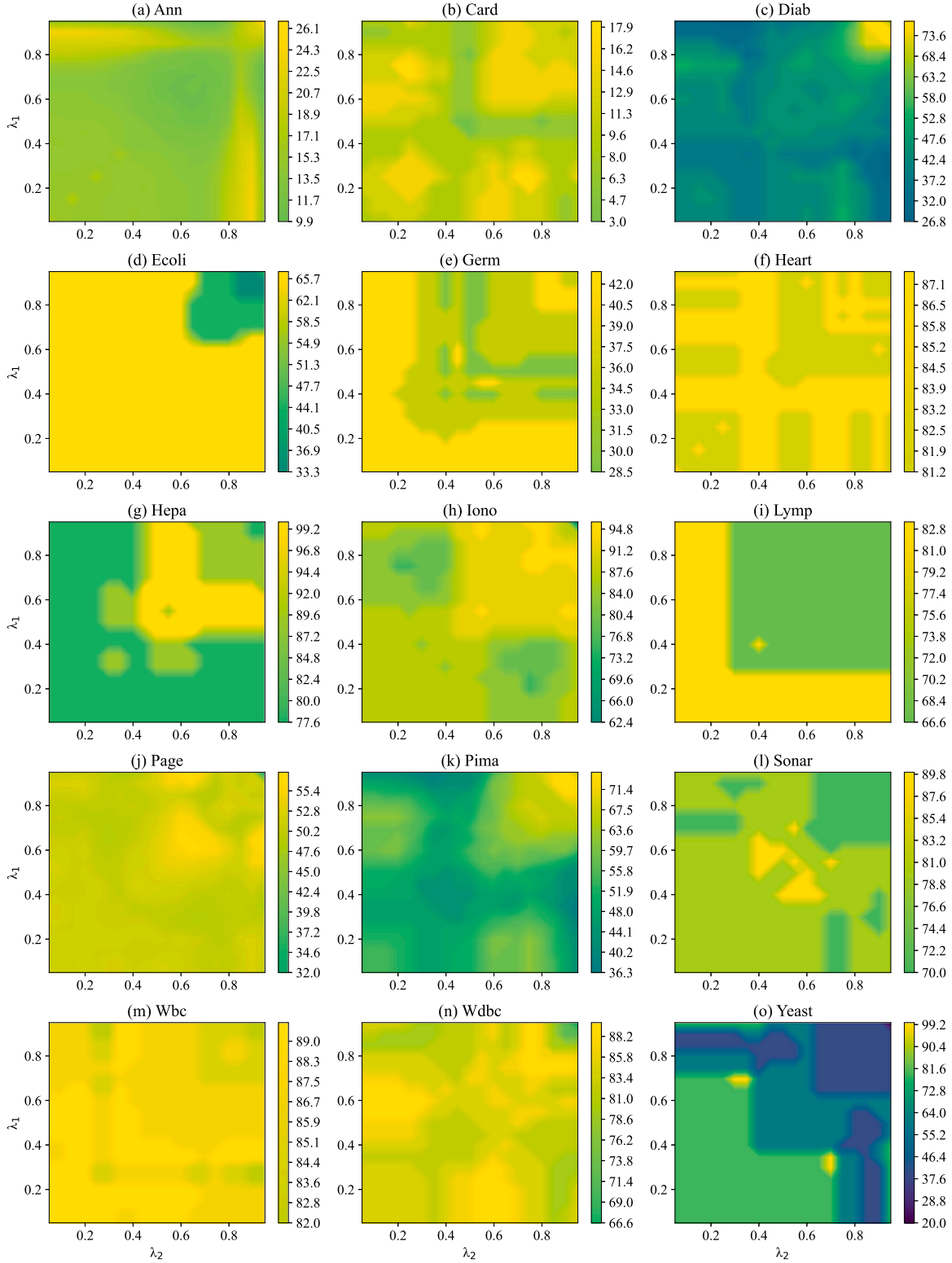

Fig. 4. $P(t)$ of MFIOD w.r.t. two-scale fuzzy granules ($t = |OT|$).

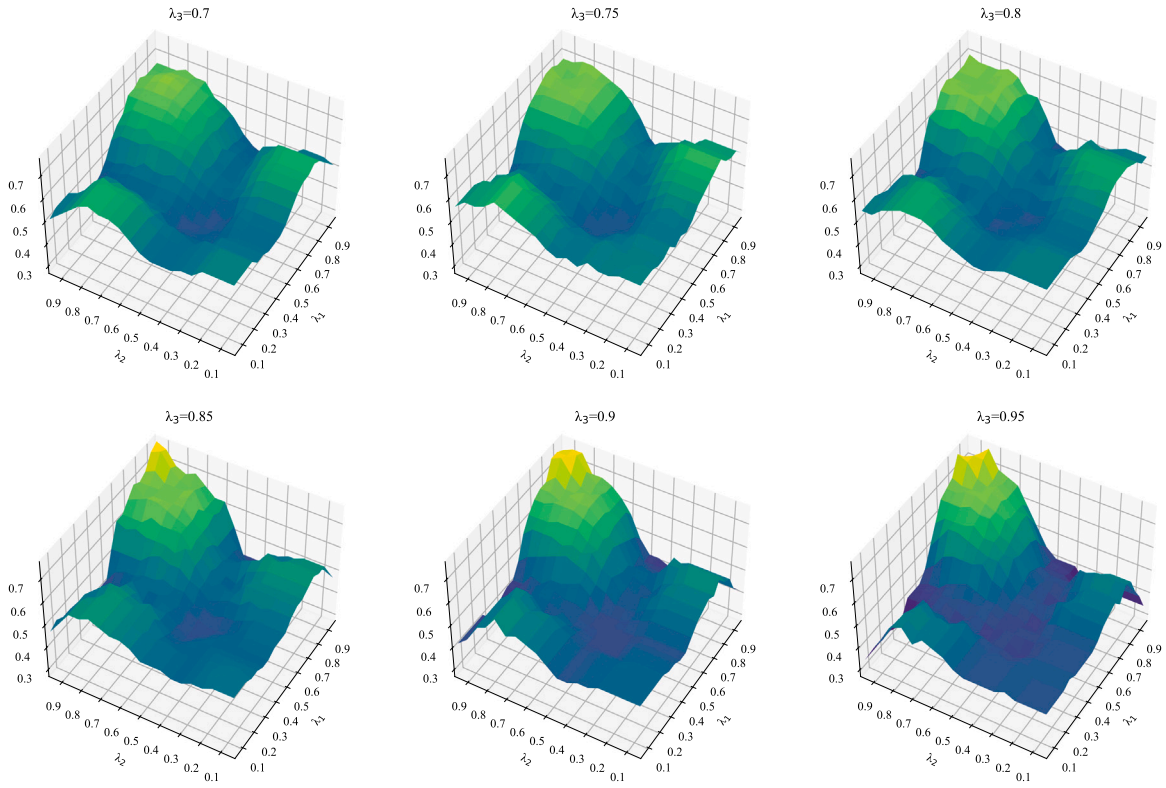

Fig. 5. $P(t)$ of MFIOD on Pima w.r.t. three-scale fuzzy granule ($t = |OT|$).

CRediT authorship contribution statement

Baiyang Chen: Conceptualization, Methodology, Formal analysis, Data curation, Investigation, Software, Writing – original draft. **Yongxiang Li:** Conceptualization, Methodology, Formal analysis, Data curation, Investigation, Software, Writing – original draft. **Dezhong Peng:** Funding acquisition, Validation, Writing – review & editing. **Hongmei Chen:** Resources, Writing – review & editing. **Zhong Yuan:** Project administration, Supervision, Writing – review & editing.

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

We have shared the link to my data/code in the manuscript.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62306196, 62376230, 61971296, 62372315, and U19A2078), Sichuan Science and Technology Program, China (2023YFQ0020, 2023YFG0033, 23ZYZYTS0077, and 2023ZHCG0016), Chengdu Science and Technology Project, China (2023-XT00-00004-GX), and the Fundamental Research Funds for the Central Universities, China (YJ202245).

References

[1] T. Pourhabibi, K.-L. Ong, B.H. Kam, Y.L. Boo, Fraud detection: A systematic literature review of graph-based anomaly detection approaches, Decis. Support Syst. 133 (2020) 113303.
[2] S. Dey, Q. Ye, S. Sampalli, A machine learning based intrusion detection scheme for data fusion in mobile clouds involving heterogeneous client networks, Inf. Fusion 49 (2019) 205–215.
[3] B. Wang, Z. Mao, Outlier detection based on Gaussian process with application to industrial processes, Appl. Soft Comput. 76 (2019) 505–516.
[4] D.M. Hawkins, Identification of Outliers, Springer, 1980.
[5] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[6] F. Jiang, Y. Sui, L. Zhou, A relative decision entropy-based feature selection approach, Pattern Recognit. 48 (7) (2015) 2151–2163.
[7] C. Wang, Y. Qian, W. Ding, X. Fan, Feature selection with fuzzy-rough minimum classification error criterion, IEEE Trans. Fuzzy Syst. 30 (8) (2022) 2930–2942.
[8] B. Sang, W. Xu, H. Chen, T. Li, Active anti-noise fuzzy dominance rough feature selection using adaptive K-nearest neighbors, IEEE Trans. Fuzzy Syst. (2023) 1–15.
[9] Z. Yuan, H.M. Chen, C. Luo, D.Z. Peng, MFGAD: Multi-fuzzy granules anomaly detection, Inf. Fusion 95 (2023) 17–25.
[10] X. Zhang, Z. Yuan, D. Miao, Outlier detection using three-way neighborhood characteristic regions and corresponding fusion measurement, IEEE Trans. Knowl. Data Eng. (2023).
[11] P. Zhang, T. Li, G. Wang, D. Wang, P. Lai, F. Zhang, A multi-source information fusion model for outlier detection, Inf. Fusion 93 (2023) 192–208.
[12] F. Li, B.Q. Hu, A new approach of optimal scale selection to multi-scale decision tables, Inform. Sci. 381 (2017) 193–208.
[13] W.-Z. Wu, Y. Leung, Theory and applications of granular labelled partitions in multi-scale decision tables, Inform. Sci. 181 (18) (2011) 3878–3897.
[14] Q. Zhang, Y. Cheng, F. Zhao, G. Wang, S. Xia, Optimal scale combination selection integrating three-way decision with hasse diagram, IEEE Trans. Neural Netw. Learn. Syst. 33 (8) (2022) 3675–3689.
[15] L. Sun, S. Si, W. Ding, X. Wang, J. Xu, TFSFB: Two-stage feature selection via fusing fuzzy multi-neighborhood rough set with binary whale optimization for imbalanced data, Inf. Fusion 95 (2023) 91–108.
[16] W.-Z. Wu, Y. Qian, T.-J. Li, S.-M. Gu, On rule acquisition in incomplete multi-scale decision tables, Inform. Sci. 378 (2017) 282–302.
[17] Y. Cheng, Q. Zhang, G. Wang, B.Q. Hu, Optimal scale selection and attribute reduction in multi-scale decision tables based on three-way decision, Inform. Sci. 541 (2020) 36–59.

[18] X. Yang, L.J. Latecki, D. Pokrajac, Outlier detection with globally optimal exemplar-based GMM, in: Proceedings of SIAM International Conference on Data Mining, SDM, 2009, pp. 145–154.
[19] M. Goldstein, A. Dengel, Histogram-based outlier score (HBOS): A fast unsupervised anomaly detection algorithm, in: KI 2012: Advances in Artificial Intelligence, vol. 9, 2012.
[20] Z. Li, Y. Zhao, X. Hu, N. Botta, C. Ionescu, G. Chen, ECOD: Unsupervised outlier detection using empirical cumulative distribution functions, IEEE Trans. Knowl. Data Eng. 35 (12) (2022) 12181–12193.
[21] F.T. Liu, K.M. Ting, Z.-H. Zhou, Isolation forest, in: 8th IEEE International Conference on Data Mining, 2008, pp. 413–422.
[22] K. Zhang, M. Hutter, H. Jin, A new local distance-based outlier detection approach for scattered real-world data, in: Pacific-Asia Conference on Knowledge Discovery and Data Mining, PAKDD, 2009, pp. 813–822.
[23] F. Angiulli, S. Basta, S. Lodi, C. Sartori, Reducing distance computations for distance-based outliers, Expert Syst. Appl. 147 (2020) 113215.
[24] M.M. Breunig, H.P. Kriegel, R.T. Ng, J. Sander, LOF: Identifying density-based local outliers, Acm Sigmod Rec. 29 (2) (2000) 93–104.
[25] 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, CIKM, 2009, pp. 1649–1652.
[26] X. Zhao, J. Liang, F. Cao, A simple and effective outlier detection algorithm for categorical data, Int. J. Mach. Learn. Cybern. 5 (3) (2014) 469–477.
[27] J. Huang, Q. Zhu, L. Yang, J. Feng, A non-parameter outlier detection algorithm based on natural neighbor, Knowl.-Based Syst. 92 (2016) 71–77.
[28] W. Hu, J. Gao, B. Li, O. Wu, J. Du, S. Maybank, Anomaly detection using local kernel density estimation and context-based regression, IEEE Trans. Knowl. Data Eng. 32 (2) (2020) 218–233.
[29] K. Li, X. Gao, S. Fu, X. Diao, P. Ye, B. Xue, J. Yu, Z. Huang, Robust outlier detection based on the changing rate of directed density ratio, Expert Syst. Appl. 207 (2022) 117988.
[30] Z.Y. He, X.F. Xu, S.C. Deng, Discovering cluster-based local outliers, Pattern Recognit. Lett. 24 (9–10) (2003) 1641–1650.
[31] L. Duan, L. Xu, Y. Liu, J. Lee, Cluster-based outlier detection, Ann. Oper. Res. 168 (1) (2009) 151–168.
[32] J. Huang, Q. Zhu, L. Yang, D. Cheng, Q. Wu, A novel outlier cluster detection algorithm without top-n parameter, Knowl.-Based Syst. 121 (2017) 32–40.
[33] 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, 2008, pp. 283–292.
[34] 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.
[35] F. Jiang, Y.M. Chen, Outlier detection based on granular computing and rough set theory, Appl. Intell. 42 (2) (2015) 303–322.
[36] M. Singh, R. Pamula, An outlier detection approach in large-scale data stream using rough set, Neural Comput. Appl. 32 (13) (2020) 9113–9127.
[37] Y.M. Chen, D.Q. Miao, H.Y. Zhang, Neighborhood outlier detection, Expert Syst. Appl. 37 (12) (2010) 8745–8749.
[38] Z. Yuan, H. Chen, T. Li, B. Sang, S. Wang, Outlier detection based on fuzzy rough granules in mixed attribute data, IEEE Trans. Cybern. 52 (8) (2022) 8399–8412.
[39] Z. Yuan, B. Chen, J. Liu, H. Chen, D. Peng, P. Li, Anomaly detection based on weighted fuzzy-rough density, Appl. Soft Comput. 134 (2023) 109995.
[40] Y. Mi, Z. Wang, H. Liu, Y. Qu, G. Yu, Y. Shi, Divide and conquer: A granular concept-cognitive computing system for dynamic classification decision making, European J. Oper. Res. 308 (1) (2023) 255–273.
[41] 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. (2021) 107353.
[42] Q. Hu, J. Liu, D. Yu, Mixed feature selection based on granulation and approximation, Knowl.-Based Syst. 21 (4) (2008) 294–304.
[43] C. Wang, Y. Huang, M. Shao, X. Fan, Fuzzy rough set-based attribute reduction using distance measures, Knowl.-Based Syst. 164 (2019) 205–212.
[44] 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.
[45] Y. Almardeny, N. Boujnah, F. Cleary, A novel outlier detection method for multivariate data, IEEE Trans. Knowl. Data Eng. 34 (9) (2022) 4052–4062.
[46] J. Dai, X. Zou, Y. Qian, X. Wang, Multifuzzy β-covering approximation spaces and their information measures, IEEE Trans. Fuzzy Syst. 31 (3) (2023) 955–969.