Information Sciences 677 (2024) 120926

Contents lists available at ScienceDirect

Information Sciences

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

Feature selection based on contradictory state sequence for multi-scale interval valued decision table

Xiaoyan Zhang *, Zihan Feng

College of Artificial Intelligence, Southwest University, Chongqing, 400715, PR China

ARTICLE INFO

Keywords:
Contradictory state sequence
Feature selection
Interval valued
Multi-scale decision table
Optimal scale selection

ABSTRACT

In the context of selecting features for multi-scale interval valued decision table ($MIVDT$), conventional research approaches encounter difficulties including elevated data complexity, diminished computational efficiency, and limited model generalization capacity. To overcome these difficulties, feature selection methods based on contradictory state sequence (CSS) and fuzzy contradictory state sequence ($FCSS$) are proposed in this paper. Initially, $MIVDT$ is established. According to the characteristics of interval value, a more thorough and impartial metric for assessing the inclusion degree is suggested, along with similarity relation based on the metric. Subsequently, the introduction of the first contradictory object allows for the delineation of contradictory state and fuzzy contradictory state, which serve to characterize the consistency of $MIVDT$. These two concepts are proposed to enhance the accuracy and efficiency of capturing key information in intricate data. Additionally, rapid feature selection algorithms are suggested, which rely on CSS and $FCSS$. In contrast to conventional feature selection approaches, the algorithms introduced in this study demonstrate superior computational efficiency and enhanced generalization capabilities when confronted with intricate datasets. In twelve open source datasets, the upper limit for the quantity of objects is 110341, and the average accuracy values of experiments under two parameter combinations are 99.54% and 97.08% respectively. The results of experiments demonstrate that the algorithms introduced are capable of efficiently identifying a novel feature subset from intricate datasets within a shorter timeframe.

1. Introduction

In the era of information technology, there has been a significant proliferation of data, posing considerable difficulties for tasks such as data analysis [14,13], risk assessment [16,21], and semantic computing [29]. One notable challenge is the issue of high-dimensional data, commonly referred to as the dimensionality disaster [35], which not only escalates the demands on computer storage and processing power but also hampers the efficacy of algorithms. Research on feature selection assumes a critical role in addressing these challenges. The objective of feature selection [1] is to identify and retain the pertinent features associated with

* Corresponding author.
E-mail addresses: zxy19790915@163.com (X. Zhang), fengzihan1218@163.com (Z. Feng).

https://doi.org/10.1016/j.ins.2024.120926
Received 5 March 2024; Received in revised form 4 June 2024; Accepted 4 June 2024

Available online 10 June 2024

the learning objective while discarding irrelevant and redundant features within extensive datasets. By employing efficient feature selection techniques, the dimensional complexity of data can be notably decreased, thereby alleviating the computational burden on computer systems. Consequently, feature selection holds significant importance across various domains including machine learning [15,28,9], approximate inference [22], concept-cognitive learning [8,31,36,5] and data mining [6,7,20].

Feature selection plays a crucial role in data analysis by enhancing model performance and efficiency, as well as improving model interpretability to facilitate a deeper understanding of the data-model relationship. Particularly when dealing with intricate and high-dimensional datasets, the judicious utilization of feature selection techniques is essential for developing a robust data analysis model. Initially, the process of feature selection typically involves exhaustive or heuristic search methods, which can be challenging to implement due to their high time and space complexities. Rough set theory is recognized as a valuable tool and approach for intelligent information processing, enabling the extraction of valuable knowledge and insights from vast, intricate datasets. Several researchers have examined the utilization of rough set theory within the domain of feature selection, yielding promising research outcomes [34,10,19,37]. Nevertheless, it is indisputable that there remain several enhancements to be implemented in this area. Primarily, the certain tools and indicators for feature selection are intricate. Despite their unique advantages, such as comprehensiveness and accuracy, they are inefficient and require substantial computing resources during practical application, which is not conducive to the completion of real-time tasks within the context of big data. Moreover, an increasing amount of data in practice is unable to be captured in discrete values, instead being commonly expressed as continuous or interval values within data tables. Consequently, the investigation of feature selection for datasets containing interval values holds significant importance, with numerous relevant studies [27,32,38,2] having been conducted. Some studies involve discretizing the interval data, while some predefine the distribution of interval value data. However, these approaches do not begin with an understanding of the inherent nature of interval values, resulting in potential errors and the omission of valuable information. In the following, the research objectives and main contributions of this paper will be elaborated respectively for these two main problems.

In the realm of big data challenges, the establishment of a straightforward and efficient metric for uncertain information is crucial. However, several existing approaches are constrained by practical considerations, making it arduous to achieve both high efficiency and accuracy simultaneously. As we all know, one of the objectives of quantifying uncertain information is to ascertain the consistency of the information table. If a direct assessment of consistency can be obtained, there is no need for a precise quantification of uncertain information, leading to a reduction in computational resource consumption. For instance, to develop a streamlined and effective feature selection technique, two indices requiring minimal space and time resources are introduced. These indices utilize Boolean or fuzzy values to directly evaluate the consistency level of the information table, enabling the swift and accurate identification of feature selection outcomes meeting the specified criteria. This study introduces these metrics, termed as contradictory state (CS) and fuzzy contradictory state (FCS) respectively. Subsequently, by constructing state sequence, feature selection can be achieved through a half-search method within the sequence. The thought can also be applied to the optimal scale selection of $MIVDT$. The study initially focuses on the optimal scale selection of $MIVDT$, followed by the completion of feature selection within the sub-information table at the optimal scale.

Some traditional methods for feature selection may not be directly applicable to interval value analysis, and existing certain approaches for handling interval values may result in the loss of valuable information and computational inaccuracies. Therefore, novel methodologies are necessary to effectively analyze interval values. The concept of pivot element inclusion degree (PI) is introduced as a means of assessing the similarity between interval values data by examining the normalized weighted overlap rate. This similarity measure, rooted in the overlap rate, allows for a thorough analysis of the data without the need for preconceived distributions, facilitating a more comprehensive approach to data analysis. In conclusion, this study offers the following contributions.

- First of all, the similarity between two interval values is assessed based on the normalized overlap ratio, and then PI with preference and its favorable properties are formulated. This metric enables efficient computation of similarity levels between interval values, exhibits favorable characteristics and broad applicability across various domains, thereby fostering new research avenues in related fields.
- Secondly, CS and FCS are utilized in the process of information table consistency judgment to immensely reduce computational complexity. The idea holds significant heuristic implications for the examination of uncertain data and offers a fresh perspective on the advancement of data analysis methodology.
- Finally, the feature selection algorithms for $MIVDT$ based on CS and FCS effectively reduce the computational cost associated with high-dimensional data processing, and enhance the processing speed and efficiency of large-scale datasets. Empirical experiments conducted on real datasets demonstrate the efficacy and practicality of the proposed methods, thereby establishing a reliable empirical foundation for subsequent related research.

The remaining paper is organized as follows. In section 2, a more comprehensive account of the related work is provided. The interval valued information system ($IVIS$) and $MIVDT$ are elucidated in detail in section 3, including PI, which is proposed according to the characteristics of interval value, and the similarity relation based on PI. Additionally, CS and FCS, as well as two optimal scale selection algorithms are elaborated in section 4. After that, two feature selection algorithms for $MIVDT$ on account of CSS and $FCSS$ are presented in section 5. Then a good deal of numerical experiments are carried out in section 6, and the efficiency of the aforementioned methods is evaluated. Ultimately, the consequences and conclusions of this paper are condensed, and the latent research directions are manifested.

2. Related works

In the introduction, the information of different studies were omitted, they are described in detail in this section.

As scientific and technological advancements progress, the variety and volume of data sources are increasing significantly. An increasing amount of data within information systems takes the form of interval values, and a single value within the interval range cannot adequately represent the entire interval. Therefore, the uncertainty associated with measuring interval data holds significant practical importance. The study of interval value data has been extensively explored by numerous researchers. A novel information entropy [3] measure was developed by Chen et al. to assess uncertainty in incomplete interval value data. This type of information entropy provides a more objective representation of the pertinent information within the interval value data, thereby enhancing the accuracy of the model. Drawing upon statistical distribution and KL divergence principles, a metric for assessing the similarity of interval values was formulated by Xu et al. [30]. The approach involves the transformation and filtering of the initial data to mitigate the issue of significant disruption to the model caused by boundary information within interval values, which can lead to a substantial decrease in overall accuracy. Ordering and analyzing interval values by setting a dominance relation on $IVIS$ was proposed by Xu et al. [32]. A new similarity degree δ_{ij}^{a} [38] was proposed by Zhang et al. by measuring the length ratio of the intersection of interval values to the union. Further interval value treatments can be found in reference. In conclusion, it is evident that certain methods lack comprehensiveness, and their incompleteness may result in the loss of valuable information. This study presents PI as a metric for assessing the resemblance between two interval values based on the normalized overlap ratio, exhibiting favorable characteristics and broad applicability across various domains.

Furthermore, within the realm of feature selection, there exist numerous notable studies [37,27,25,23,22]. By introducing different weights into neighborhood relations, a new feature selection based weighted neighborhood rough set ($IVWNRS$) model for $IVIS$ was delineated by Zhang et al. [37]. This method considers the different attribute weights in $IVIS$, and proposes the weighted neighborhood relation to solve the contradiction between the dependency degree of attribute subset and the classification ability. Weighted dominant neighborhood rough set ($WDNRS$) by assigning varying weights to conditional attributes was introduced by Pan et al. [19]. Also a matrix representation of conditional entropy and a updating method to assess attribute significance were presented. Ultimately, a heuristic method for selecting features and an associated incremental mechanism were suggested for scenarios involving an increase in objects. Attribute reduction based on distance granulation and conditional entropy in incomplete interval valued decision system was delineated by Chen et al. [2]. The method elevates the distance measure and similarity degree through range completion and statistical enhancement, and improves the conditional entropy by utilizing coverage-credibility instead of credibility, thereby boosting the classification performance of the algorithm. An improved binary meerkat optimization algorithm for efficient feature selection of supervised learning classification was proposed by Hussien et al. [12]. This approach improves feature exploration and utilization by integrating the periodic mode boundary handling ($PMBH$) strategy with the local search (LS) process, thereby enhancing classification accuracy. Novel thresholds for interval dominance degree (IDD) and interval overlap degree (IOD) among interval values were introduced by Li et al. [18]. Subsequently, the interval valued dominance relation is established utilizing these parameters, followed by an examination of the interval valued dominance rough set approach ($IV-DRSA$) and its associated characteristics. Lastly, a feature selection guideline based on IDD derived from $IV-DRSA$ is proposed. Various approaches have demonstrated notable success to a certain degree. Nevertheless, certain existing approaches are constrained by practical considerations, such as heightened computational demands and susceptibility to alterations in data distribution. Presently, achieving concise and efficient methods for feature selection remains a challenging task. For additional feature selection algorithms and concepts, please refer to [24,17,10,4,1].

3. Interval valued information system and multi-scale interval valued decision table

In this section, the cardinal concept of $IVIS$ is retrospected. Furthermore, the introduction of $MIVDT$ is accompanied by the establishment of surjective relationships between neighboring scales. This serves to facilitate the maintenance of monotonicity in similarity relation and similarity class across various scales and condition attribute subsets. Simultaneously, to assess the resemblance of interval values established within the domain of real numbers, the concept of PI is introduced based on the intersection of interval values, along with the elucidation of four critical properties of PI. Subsequently, a consistency definition of $MIVDT$ is presented.

3.1. Interval valued information system

Definition 1. [32] An $IVIS$ can be visually revealed by a tuple $S=(O,C)$. $O=\{o_1,o_2,...,o_n\}$ is a non-empty limited set of objects, which is considered as the universe of discourse; $C=\{c_1,c_2,...,c_m\}$ is a non-empty bounded set of attributes, where c_j can be regarded as a mapping, $c_j: O\to W(V_j)$, for all $c_j\in C$, i.e. $c_j(o_i)=[a_j^L(o_i),a_j^U(o_i)]$, $o_i\in O$, where $a_j^L(o_i)\le a_j^U(o_i)$, $a_j^L(o_i)\in V_j$, $a_j^U(o_i)\in V_j$, and V_j is the domain of attribute c_j, $W(V_j)$ is the set of all interval values over V_j.

3.2. Multi-scale interval valued decision table

Definition 2. [26] A MIS can be pictorially demonstrated by a tuple $S=(O,C)$. $O=\{o_1,o_2,...,o_n\}$ is a non-empty restricted set of objects, which is referred to as the universe of discourse; $C=\{c_1,c_2,...,c_m\}$ is a non-empty finite set of conditional attributes, each

Table 1
A multi-scale interval valued decision table.

O	c_1		c_2		d
	c_1^1	c_1^2	c_2^1	c_2^2	
o_1	[96, 96]	[9, 10]	[45, 83]	[4, 9]	1
o_2	[30, 36]	[3, 4]	[1, 19]	[0, 2]	1
o_3	[20, 66]	[2, 7]	[32, 55]	[3, 6]	2
o_4	[18, 52]	[1, 6]	[55, 89]	[5, 9]	2
o_5	[24, 35]	[2, 4]	[5, 36]	[0, 4]	1
o_6	[15, 59]	[1, 6]	[1, 72]	[0, 8]	1
o_7	[47, 50]	[4, 5]	[22, 77]	[2, 8]	2
o_8	[41, 98]	[4, 10]	[44, 85]	[4, 9]	2

conditional attribute has L scales. Where c_j^l can be deemed as a mapping, $c_j^l : O \to V_j^l$, for all $c_j \in C$, where V_j^l is the domain of conditional attribute c_j on the $l-th$ scale. A MIS can be expanded to the following form.

$$S = (O, C) = (O, \{c_j^l | j = 1, 2, ..., m; l = 1, 2, ..., L\}). \tag{1}$$

Definition 3. A $MIVDT$ can be pictorially held up by a tuple $S = (O, C \cup \{d\})$. $O = \{o_1, o_2, ..., o_n\}$ is a non-empty restricted set of objects, which is referred to as the universe of discourse; $C = \{c_1, c_2, ..., c_m\}$ is a non-empty finite set of conditional attributes, each conditional attribute has L scales. Where c_j^l can be deemed as a mapping, $c_j^l : O \to W(V_j^l)$, i.e. $c_j^l(o_i) = [a_j^{l,L}(o_i), a_j^{l,U}(o_i)]$, $o_i \in O$, where $a_j^{l,L}(o_i) \le a_j^{l,U}(o_i)$, $a_j^{l,L}(o_i) \in V_j^l$, $a_j^{l,U}(o_i) \in V_j^l$, $W(V_j^l)$ is the set of all interval values over V_j^l, for all $c_j \in C$, where V_j^l is the domain of conditional attribute c_j on the $l-th$ scale. In addition, $d \notin \{c_j^l | j = 1, 2, ..., m; l = 1, 2, ..., L\}$ the decision such that $d : O \to V_d$, where V_d is the domain of decision attribute d with a single scale. A $MIVDT$ can be expanded to the following form.

$$S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, ..., m; l = 1, 2, ..., L\} \cup \{d\}),\ c_j^l = [a_j^{l,L}, a_j^{l,U}]. \tag{2}$$

It is crucial to recognize that the interval values studied in this research exhibit slight variations from those delineated in existing literature, thereby warranting a re-definition of this concept within the current framework. Let $e^L, e^U \in R$, and $e^L \le e^U$, then an interval value over the field of real numbers can be defined as follows.

$$E = [e^L, e^U] = \{e \in R | e^L \le e \le e^U\}. \tag{3}$$

Where e^L and e^U constitute the lower and upper bounds of E individually. Additionally, the empty interval value is consistently recounted with $[*, *]$, $* \notin R$. The set of all interval values on the real number field is narrated as $W(R)$. In particular, when $e^L = e^U$, the interval value E will degenerate to a real number. The study of $IVIS$ also plays a role in advancing the understanding of single-valued information systems. The interval values are restricted to non-negative values for the purpose of enhancing clarity. A $MIVDT$ is revealed in Table 1.

For a $MIVDT$ $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, ..., m; l = 1, 2, ..., L\} \cup \{d\})$, it can be disintegrated into L $sub-IVDTs$, indicated as $S^l = (O, C^l \cup \{d\}), l = 1, 2, ..., L$. In a MIS, there is a surjective relationship between adjacent scales of the identical conditional attribute. For each $l \in \{1, 2, ..., L-1\}$, $j \in \{1, 2, ..., m\}$, there exists a surjective mapping $\phi_j^{l,l+1} : V_j^l \to V_j^{l+1}$, i.e. $c_j^{l+1}(o_i) = \phi_j^{l,l+1}(c_j^l(o_i)), o_i \in O$. Where $\phi_j^{l,l+1}$ is known as scale transformation mapping. Equally, in a $MIVDT$, there are two surjective relationship between adjacent scales of the duplicate conditional attribute. For each $l \in \{1, 2, ..., L-1\}$, $j \in \{1, 2, ..., m\}$, there endures two surjective mapping, one is $\phi_j^{l,l+1} : V_j^l \to V_j^{l+1}$, i.e. $v_j^{l+1}(o_i) = \phi_j^{l,l+1}(v_j^l(o_i)), o_i \in O$, $v_j^l \in V_j^l, v_j^{l+1} \in V_j^{l+1}$. The another is $\Phi_j^{l,l+1} : W(V_j^l) \to W(V_j^{l+1})$, i.e. $c_j^{l+1}(o_i) = \Phi_j^{l,l+1}(c_j^l(o_i)) = \Phi_j^{l,l+1}([a_j^{l,L}(o_i), a_j^{l,U}(o_i)]) = [a_j^{l+1,L}(o_i), a_j^{l+1,U}(o_i)], o_i \in O$. The two mappings have the following associations.

$$\Phi_j^{l,l+1}([a_j^{l,L}(o_i), a_j^{l,U}(o_i)]) = [\phi_j^{l,l+1,L}(a_j^{l,L}(o_i)), \phi_j^{l,l+1,U}(a_j^{l,U}(o_i))]. \tag{4}$$

Where $\phi_j^{l,l+1}$ is regarded as scale transformation mapping, $\phi_j^{l,l+1,L}$ and $\phi_j^{l,l+1,U}$ can be the deviating or assorted, and possess miscellaneous pragmatic implications in contrasting realistic circumstances. $\Phi_j^{l,l+1}$ is considered as domain transformation mapping. In Table 1, $O = \{o_1, o_2, ..., o_8\}$ stands for eight graduate students. c_1 and c_2 represent the range of scores achieved by eight students across 10 tests in two distinct courses, which are *Logic in Computer Science* and *Applied Stochastic Processes*. Scale 1 reveals the grades obtained through written assessments, while scale 2 delineates the evaluations provided by teachers. The scale transformation mapping and domain transformation mapping between two scales can be defined as follows, $j \in \{1, 2\}$, $i \in \{1, 2, 3, 4, 5, 6, 7, 8\}$.

$$\Phi_j^{1,2}([a_j^{1,L}(o_i), a_j^{1,U}(o_i)]) = [\phi_j^{1,2,L}(a_j^{1,L}(o_i)), \phi_j^{1,2,U}(a_j^{1,U}(o_i))] = [a_j^{2,L}(o_i), a_j^{2,U}(o_i)]. \tag{5}$$

$$\phi_j^{1,2,L}(v)=\begin{cases}0 & 0\le v<10\\ 1 & 10\le v<20\\ 2 & 20\le v<30\\ 3 & 30\le v<40\\ 4 & 40\le v<50\\ 5 & 50\le v<60\\ 6 & 60\le v<70\\ 7 & 70\le v<80\\ 8 & 80\le v<90\\ 9 & 90\le v<100\\ 10 & v=100\end{cases},\quad \phi_j^{1,2,U}(v)=\begin{cases}0 & v=0\\ 1 & 0<v\le 10\\ 2 & 10<v\le 20\\ 3 & 20<v\le 30\\ 4 & 30<v\le 40\\ 5 & 40<v\le 50\\ 6 & 50<v\le 60\\ 7 & 60<v\le 70\\ 8 & 70<v\le 80\\ 9 & 80<v\le 90\\ 10 & 90<v\le 100\end{cases}. \tag{6}$$

A binary relation on objects contained by d is interpreted as follows.

$$R_d=\{(o,p)\in O\times O|d(o)=d(p)\}. \tag{7}$$

Relation R_d consists of a partition O/R_d on O, written as $O/R_d=\{R_d(o)|o\in O\}=\{D_1,D_2,...,D_t\}$, where $R_d(o)=\{p|(o,p)\in R_d, p\in O\}$, t is the total number of decision classes. Correspondingly, in a $MIVDT$, there are similarity relation and similarity class generated based on conditional attributes. To define the similarity relation, an index is required to quantify the level of similarity between two entities. Evidently, a higher percentage of overlap within two interval values results in a more pronounced similarity of the interval values. Simultaneously, to achieve a more impartial and accurate assessment of the similarity value, the degree of overlap between two interval values is individually quantified by dividing the length of their intersection by the length of the interval value. The length is determined by the difference between the endpoints. The resulting two overlap degrees are standardized and combined to ensure that their similarity falls within $[0,1]$. This process establishes the definition of PI.

Definition 4. For two non-empty interval values $E=[e^L,e^U]$ and $F=[f^L,f^U]$ over the field of real numbers, when $e^L\le f^L$ and $\max\{e^L,f^L\}\le\min\{e^U,f^U\}$, the intersection operation of two interval values can be defined as follows.

$$E\cap F=F\cap E=[\max\{e^L,f^L\},\min\{e^U,f^U\}]=\{g\in R|g\in E,g\in F\}. \tag{8}$$

When $E=[*,*]$ or $F=[*,*]$, $E\cap F=[*,*]$. When $\max\{e^L,f^L\}>\min\{e^U,f^U\}$, this manifests that E and F don't intersect, then $E\cap F=[*,*]$. When $f^L\le e^L$, a homogeneous conclusion can be secured, which will not be repeated.

Definition 5. Supposing two interval values $E=[e^L,e^U]$ and $F=[f^L,f^U]$ bounded over the field of real numbers, and $E\cap F=[g^L,g^U]$, PI of E w.r.t. F can be defined as follows.

$$PI(E,F)=\zeta_1\frac{g^U-g^L}{e^U-e^L}+\zeta_2\frac{g^U-g^L}{f^U-f^L},\ \zeta_1+\zeta_2=1. \tag{9}$$

Proposition 1. *Below are four characteristics of Equation (9).*

- *If* $E=F=[e,e], e\in R$, *then* $PI(E,F)=1$.
- *If* $E=[e,e]$ *and* $F=[f,f]$, $e\ne f$, $e,f\in R$, *then* $PI(E,F)=0$.
- *When* $*$ *appears,* $PI(E,F)=0$.
- $0\le PI(E,F)\le 1$, *and when* $\zeta_1=\zeta_2=\frac{1}{2}$, $PI(E,F)=PI(F,E)$.

The concrete process of calculating PI is exhibited in Fig. 1. It is undeniable that PI holds greater practical significance. For example, when analyzing the structure of a social network, it is essential to carefully scrutinize the social connections of each person. Only by taking this thorough approach can we generate various categorizations of interpersonal relationships, leading to a better understanding of individuals' roles within the group. In Fig. 2, an asymmetric directed chord diagram is applied to demonstrate multiple social relationships within a randomly generated social relationship group. The directionality in a relationship is denoted by the quantity of lines. Different nodes represent a variety of objects, with the color of these object nodes indicating different categories of objects, and the size of object nodes representing the complexity of social relationships. It is commonly acknowledged that certain social relationships, such as admiration, hostility, and affection, frequently exhibit a unilateral nature, requiring a focus on a particular entity for their formation. Therefore, the chord diagram in Fig. 2 is directed. These realistic factors coincide with the setting of PI with preference. According to abundant exploration and distinguishing analysis, ζ_1,ζ_2 in PI can be properly configured to make PI applicable to a variety of discrepant situations.

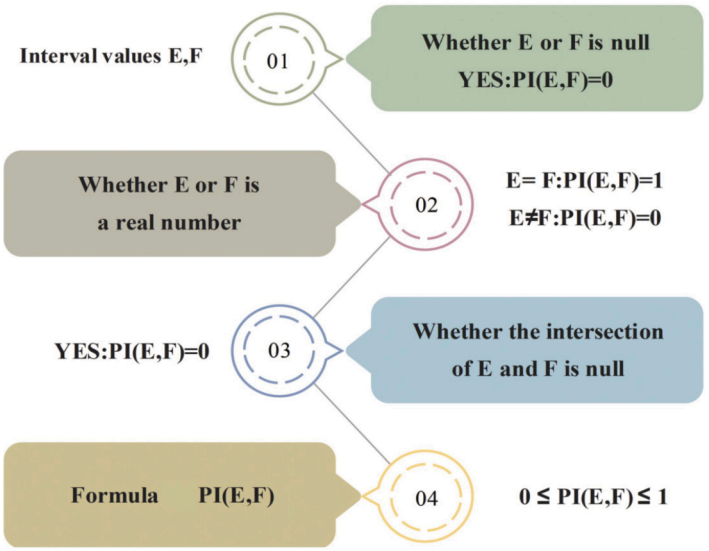

Fig. 1. Flow chart for calculating PI.

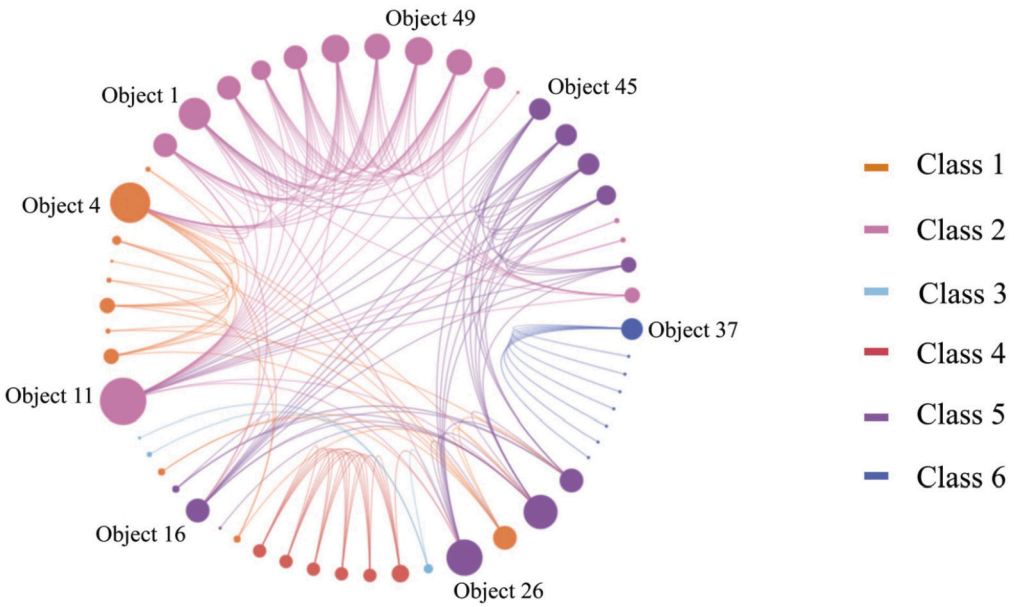

Fig. 2. Asymmetric directed social relation chord diagram.

The similarity relation based on PI is given below, which can conceive object taxonomies on account of conditional attributes at different scales.

Definition 6. Let $S = (O, C \cup \{d\}) = (O, \{c_j^l | j = 1, 2, ..., m; l = 1, 2, ..., L\} \cup \{d\})$ be a $MIVDT$, $\varsigma^l \in [0, 1]$, besides $B \subseteq C$, for $l \in \{1, 2, ..., L\}$, a similarity relation w.r.t. ς^l and B^l is portrayed as follows.

$$R_{B^l}^{\varsigma^l} = \{< o, p > \in O \times O | PI(c^l(o), c^l(p)) \geq \varsigma^l, \forall c \in B\}. \tag{10}$$

When PI takes divergent values of ζ_1 and ζ_2, maybe $PI(c^l(o), c^l(p)) \neq PI(c^l(p), c^l(o))$. For that reason, in multifarious practical applications, it is contingent that $PI(c^l(o), c^l(p)) \geq \varsigma^l$, nevertheless $PI(c^l(p), c^l(o)) < \varsigma^l$, i.e. $< o, p > \in R_{B^l}^{\varsigma^l}$, $< p, o > \notin R_{B^l}^{\varsigma^l}$ yet. The similarity class of each object $o \in O = \{o_1, o_2, o_3, ..., o_n\}$ w.r.t. B^l and ς^l is proclaimed as follows, for $l \in \{1, 2, ..., L\}$, $B \subseteq C$.

$$SC_{B^l}^{\varsigma^l}(o) = \{p | < o, p > \in R_{B^l}^{\varsigma^l}, p \in O\}. \tag{11}$$

For a $MIVDT$, PI between objects may shrink as the scale evolves into coarse. Under the circumstances, if the threshold ς in Equation (10) abides invariant, there is a possibility of smaller similarity classes as the scale grows into coarse. However, as the scale increases, the similarity classes should also increase. To address this issue, make sure that ς in Equation (10) varies as the scale transforms, specify the threshold ranges [27] for diverse scales as follows.

$$\varsigma^{l+1} \le \kappa^{l+1} \le \varsigma^{l},\ l \in \{1,2,...,L-1\},\ \kappa^{l+1} = \min\left(\min_{c\in C}\left(\min_{<o,p>\in R^{\varsigma^l}_{C^l}}(PI(c^{l+1}(o),c^{l+1}(p)))\right),\varsigma^{l}\right). \tag{12}$$

According to the proclamation of Equation (12), it can be obtained that $R^{\varsigma^l}_{C^l} \subseteq R^{\varsigma^{l+1}}_{C^{l+1}}$.

Proof. For any $<o,p>\in R^{\varsigma^l}_{C^l}, x\in C,\ PI(x^{l+1}(o),x^{l+1}(p)) \ge \kappa^{l+1} \ge \varsigma^{l+1}$. Consequently, $<o,p>\in R^{\varsigma^{l+1}}_{C^{l+1}}$ and $R^{\varsigma^l}_{C^l} \subseteq R^{\varsigma^{l+1}}_{C^{l+1}}$ can be procured.

$$PI(x^{l+1}(o),x^{l+1}(p)) \ge \min_{c\in C}\left(\min_{<o,p>\in R^{\varsigma^l}_{C^l}}(PI(c^{l+1}(o),c^{l+1}(p)))\right). \tag{13}$$

The set of similarity classes of all objects is recorded as $SC^{\varsigma^l}_{B^l} = \{SC^{\varsigma^l}_{B^l}(o_1), SC^{\varsigma^l}_{B^l}(o_2), SC^{\varsigma^l}_{B^l}(o_3),..., SC^{\varsigma^l}_{B^l}(o_n)\}$, for $l\in\{1,2,...,L\}$, $B\subseteq C$. It is conspicuous that $o_i \in SC^{\varsigma^l}_{B^l}(o_i)$, $SC^{\varsigma^l}_{B^l}(o_i)\neq\emptyset$, $\underset{o_i\in O}{\cup} SC^{\varsigma^l}_{B^l}(o_i) = O$, for all $o_i\in O$, thus $SC^{\varsigma^l}_{B^l}$ constitutes a covering on O.

Definition 7. Let $S=(O,C\cup\{d\})$ be an $IVDT$, $\varsigma\in[0,1]$, as well as $B,H\subseteq C$, if $SC^{\varsigma}_{B}(o_i)\subseteq SC^{\varsigma}_{H}(o_i)$, for all $o_i\in O$, we announce that SC^{ς}_{B} is finer than SC^{ς}_{H}, afterwards assert it as $SC^{\varsigma}_{B}\preceq SC^{\varsigma}_{H}$.

Proposition 2. *Let $S=(O,C\cup\{d\})$ be a $MIVDT$, for $l\in\{1,2,...,L-1\}$.*

$$SC^{\varsigma^l}_{C^l}(o_i)\subseteq SC^{\varsigma^{l+1}}_{C^{l+1}}(o_i),\ \forall o_i\in O,\ SC^{\varsigma^l}_{C^l}\preceq SC^{\varsigma^{l+1}}_{C^{l+1}}. \tag{14}$$

The above establishes the monotonicity of similarity class in a $MIVDT$, which indicates that, as the scale increases, the covering on the universe becomes coarse.

Definition 8. Let $S=(O,C\cup\{d\})$ be a $MIVDT$, plus $\varsigma^l\in[0,1]$, for $l\in\{1,2,...,L\}$. S is said to be consistent if $S^1=(O,C^1\cup\{d\})$ is consistent, i.e. $SC^{\varsigma^1}_{C^1}(o_i)\subseteq R_d(o_i)$, for all $o_i\in O$. In contrary fashion, S is inconsistent. Also, $S^l=(O,C^l\cup\{d\})$ is consistent, if and only if $SC^{\varsigma^l}_{C^l}(o_i)\subseteq R_d(o_i)$, for all $o_i\in O$.

To avoid repetitive evaluation of the information table consistency during feature selection, all algorithms accomplish a consistency check on universe of discourse O before execution, i.e. delete o_i, if $SC^{\varsigma^1}_{C^1}(o_i)\not\subseteq R_d(o_i)$, for all $o_i\in O$. Obviously, as the scale becomes rough, the consistency of $S^l=(O,C^l\cup\{d\})$ should become worse and worse. However, if thresholds ς are not addressed, it is likely that $S^l=(O,C^l\cup\{d\})$ becomes more consistent as the scale gets rougher. This is sufficient to explain the necessity of threshold ς^l processing.

Proposition 3. *Let $S=(O,C\cup\{d\})$ be an $IVDT$, $\varsigma\in[0,1]$, in addition to $B,H\subseteq C$, if $H\subseteq B$, it follows that $SC^{\varsigma}_{B}\preceq SC^{\varsigma}_{H}$ and $R^{\varsigma}_{B}\subseteq R^{\varsigma}_{H}$.*

4. Optimal scale selection for multi-scale interval valued decision table

Initially, it is imperative to elucidate the concept of optimal scale and feature selection within the context of this study. The optimal scale refers to the maximum scale required to maintain consistency in the sub-information table S^l. Similarly, the objective of feature selection is to identify the smallest subset of attributes that ensures consistency in the optimal scale sub-information table. In the process of judging the consistency of $MIVDT$, two contradiction detection indexes are proposed, according to which CS and FCS are given to measure the degree of consistency. The distinction lies in the fact that CS can only assess consistency based on Boolean values, whereas FCS manifests the degree of inconsistency of S^l through fuzzy value. Then the optimal scale selection of $MIVDT$ is solved by using CSS and $FCSS$, which makes preparation for the subsequent feature selection. The main model of the full text is presented in Fig. 3, serving as a convenient reference for subsequent elucidation.

4.1. Contradictory state and fuzzy contradictory state

Definition 9. In a $MIVDT$ $S=(O,C\cup\{d\})$, $O=\{o_1,o_2,...,o_n\}$. When $S^l=(O,C^l\cup\{d\})$, we judge whether $o_i\in O$ satisfy $SC^{\varsigma^l}_{C^l}(o_i)\subseteq R_d(o_i)$ in line with the order of subscripts. The first object that does not gratify $SC^{\varsigma^l}_{C^l}(o_i)\subseteq R_d(o_i)$ is called the first

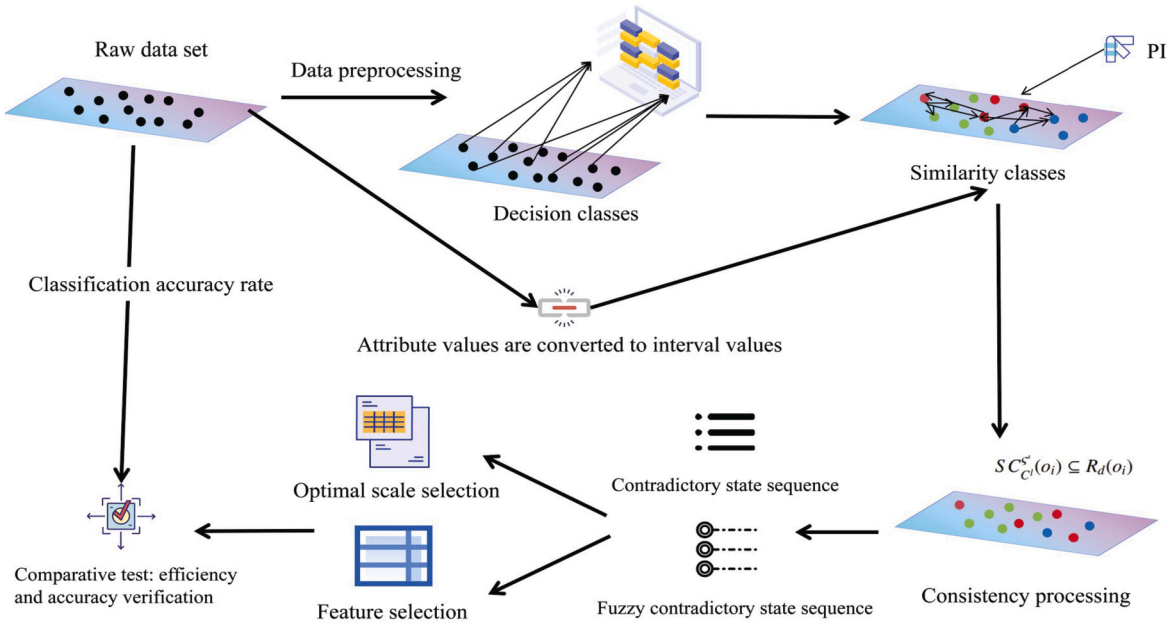

Fig. 3. Model flow chart.

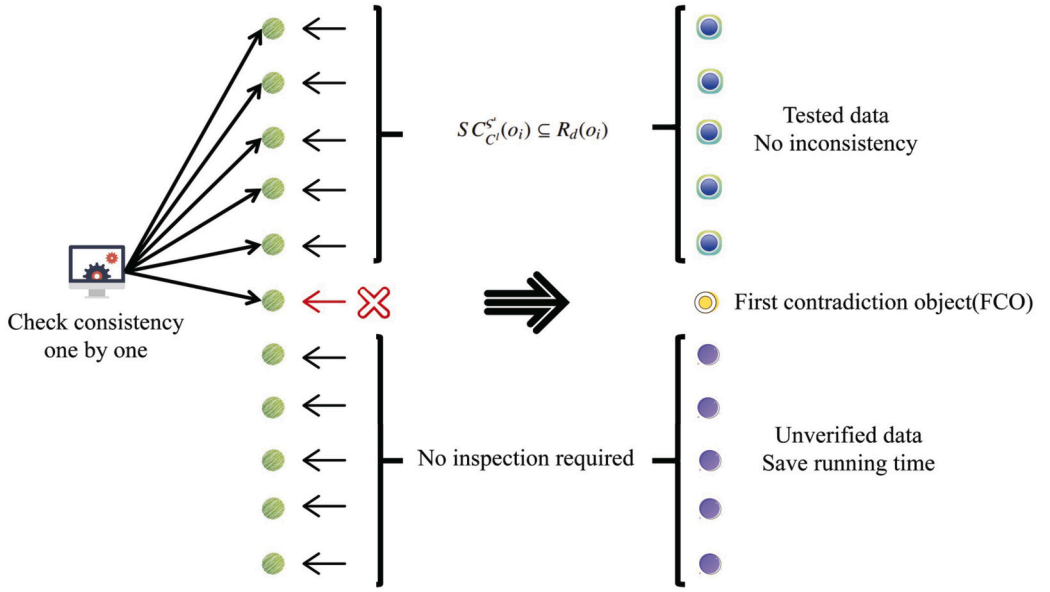

Fig. 4. The process of finding the FCO and FCI.

contradictory object (FCO) w.r.t. $S^l = (O, C^l \cup \{d\})$, denoted by $FCO(S^l) = o_i$. And i is the first contradictory index (FCI) w.r.t. $S^l = (O, C^l \cup \{d\})$, recorded by $FCI(S^l) = i$. If $S^l = (O, C^l \cup \{d\})$ is consistent, then $FCI(S^l) = n + 1$.

The process of finding the FCO and FCI under an information subtable is illustrated in Fig. 4.

Definition 10. In a $MIVDT$ $S = (O, C \cup \{d\})$, $l \in \{1, 2, ..., L\}$, $S^l = (O, C^l \cup \{d\})$, the form of CS is defined as follows.

$$CS(S^l) = CS(O, C^l \cup \{d\}) = \begin{cases} 0, & FCI(S^l) = n + 1 \\ 1, & FCI(S^l) \neq n + 1 \end{cases}. \tag{15}$$

It is unambiguous that $CS(S^l)$ is a Boolean value, either 0 or 1. When $S^l = (O, C^l \cup \{d\})$ is consistent, i.e. $SC^{\varsigma^l}_{C^l}(o_i) \subseteq R_d(o_i)$, for all $o_i \in O$, $CS(S^l) = 0$, otherwise $CS(S^l) = 1$. The result of $CS(S^l)$ can be obtained by checking whether the condition

$SC_{C^l}^{\varsigma^l}(o_i) \subseteq R_d(o_i)$ is true in turn. In this process, we can quit checking only when we discover $FCO(S^l) = o_i$, $o_i \in O$, $i < n+1$, then $CS(S^l) = 1$.

Definition 11. In a $MIVDT$ $S = (O, C \cup \{d\})$, $l \in \{1, 2, ..., L\}$, $S^l = (O, C^l \cup \{d\})$, the form of FCS is defined as follows.

$$FCS(S^l) = FCS(O, C^l \cup \{d\}) = \lambda,\ \lambda \in [0, 1]. \tag{16}$$

When S^l is consistent, $FCS(S^l) = 0$. When $FCS(S^l) \neq 0$, its value indicates the degree of inconsistency in S^l. A larger $FCS(S^l)$ implies greater inconsistency in S^l and more satisfaction of $SC_{C^l}^{\varsigma^l}(o_i) \nsubseteq R_d(o_i)$ by o_i, for all $o_i \in O$. In diverse scenarios, $FCS(S^l)$ varies along with their respective meanings.

When multiple CSs are arranged together, they may show some regular changes, which is used to represent the change process of a concrete phenomenon. Similarly, $FCSs$ can also be employed. For the feature selection problem of $MIVDT$, sequences with the following formal characteristics can be stipulated. A monotonically undecreasing ordered sequence consisting of countably finite CSs is called CSS. The monotonic undecreasing ordered sequence composed of countably finite $FCSs$ is called $FCSS$. The following will introduce and apply several special sequences, such as optimal scale selection contradictory state sequence ($OSCS$), optimal scale selection fuzzy contradictory state sequence ($OSFCS$), feature selection contradictory state sequence ($FSCS$), feature selection fuzzy contradictory state sequence ($FSFCS$).

4.2. Optimal scale selection for multi-scale interval valued decision table

In a $MIVDT$ $S = (O, C \cup \{d\})$, $O = \{o_1, o_2, ..., o_n\}$, $l \in \{1, 2, ..., L\}$, $\varsigma^l \in [0, 1]$, $S^l = (O, C^l \cup \{d\})$. On account of Equation (14), when $SC_{C^l}^{\varsigma^l}(o_i) \nsubseteq R_d(o_i)$, $SC_{C^{l+1}}^{\varsigma^{l+1}}(o_i) \nsubseteq R_d(o_i)$, for all $o_i \in O$. This reveals that as the scale becomes rough, the inconsistency don't get weaker. If l is taken from 1 to L, the corresponding $S^l = (O, C^l \cup \{d\})$ generates $CS(S^l)$ and $FCS(S^l)$, then forming $OSCS$ and $OSFCS$ respectively.

$$OSCS = [CS(S^1), CS(S^2), ..., CS(S^L)]. \tag{17}$$

The optimal scale of S is determined by the final sequence of 0, which commences at 1. The obstacle of seeking the optimal scale is then transformed into the problem of locating a certain number in a sequential array. When employing the sequential search approach, it may yield improved results with a small number of scales. However, when dealing with large data sets at multiple scales, the half-search method may be more effective. Therefore, when utilizing $OSCS$ to find the optimal scale of S, the more efficient half-search method is employed.

In the course of executing four algorithms, the similarity class of individual objects is produced and handled in isolation, as opposed to simultaneously generating and storing all similarity classes to mitigate the risk of memory exhaustion. Moreover, as each object is exclusively associated with a single decision class, the comparison of the similarity class of the object is required only against its specific decision class, rather than against all decision classes. The time complexity evaluation of the four algorithms is extensive, therefore they have been organized in Table 2 ~ Table 5 correspondingly.

Example 1. In Table 1, assuming $\varsigma^1 = \varsigma^2 = 0.8$, $\zeta_1 = \frac{2}{3}, \zeta_2 = \frac{1}{3}$, the decision classes and similarity classes under two scales are acquired respectively. After calculating $O/R(d) = \{\{1, 2, 5, 6\}, \{8, 3, 4, 7\}\}$, $FCI(S^1) = 9$, $FCI(S^2) = 9$, $CS(S^1) = 0$, and $CS(S^2) = 0$, in this way the optimal scale $OS = 2$.

$$SC_{C^1}^{\varsigma^1}(o_1) = \{1\} \qquad SC_{C^2}^{\varsigma^2}(o_1) = \{1\}$$
$$SC_{C^1}^{\varsigma^1}(o_2) = \{2\} \qquad SC_{C^2}^{\varsigma^2}(o_2) = \{2, 5\}$$
$$SC_{C^1}^{\varsigma^1}(o_3) = \{3\} \qquad SC_{C^2}^{\varsigma^2}(o_3) = \{3\}$$
$$SC_{C^1}^{\varsigma^1}(o_4) = \{4\} \qquad SC_{C^2}^{\varsigma^2}(o_4) = \{4\}$$
$$SC_{C^1}^{\varsigma^1}(o_5) = \{5\} \qquad SC_{C^2}^{\varsigma^2}(o_5) = \{5\}$$
$$SC_{C^1}^{\varsigma^1}(o_6) = \{6\} \qquad SC_{C^2}^{\varsigma^2}(o_6) = \{6\}$$
$$SC_{C^1}^{\varsigma^1}(o_7) = \{7\} \qquad SC_{C^2}^{\varsigma^2}(o_7) = \{7\}$$
$$SC_{C^1}^{\varsigma^1}(o_8) = \{8\} \qquad SC_{C^2}^{\varsigma^2}(o_8) = \{8\}$$

In $OSFCS$, if $FCS(S^l) = 0$, S^l is consistent. Then the optimal scale to be pursued, that is, the largest scale at which S^l can be verified for consistency. The concept of $OSFCS$ elucidates the progression from consistency to inconsistency as the scale becomes coarser, representing a gradual rather than abrupt transformation. Additionally, it is prerequisite to enumerate the growth rule of FCS in $OSFCS$. The requirement for this rule is that it must exhibit monotonicity and have a range of $[0, 1]$. To simplify the process and improve the effectiveness of Algorithm 2, a FCS growth rule based on the strength of FCI is defined. This

Algorithm 1: Optimal scale selection based on optimal scale selection contradictory state sequence.

Input: A $MIVDT\ S=(O,C\cup\{d\})$, $O=\{o_1,o_2,...,o_n\}$. L threshold values ς^l, L scales $l\in\{1,2,...,L\}$.
Output: Optimal scale OS.

Calculating $O/R_d=\{R_d(o)|o\in O\}=\{D_1,D_2,...,D_t\}$
/*Start finding optimal scale by the half-search.*/
$left=1, right=L$
while $left\le right$ **do**
  $mid=(left+right)//2$
  /*Calculating $FCI(S^{mid})$.*/
  $FCI(S^{mid})$=n+1
  /*Verification from the first object to the final object.*/
  for $i=1\rightarrow n$ **do**
    Calculating $SC^{\varsigma^{mid}}_{C^{mid}}(o_i)$
    if $SC^{\varsigma^{mid}}_{C^{mid}}(o_i)\not\subseteq R_d(o_i)$ **then**
      $FCI(S^{mid})$=i
      break
    end if
  end for
  /*$FCI(S^{mid})$ is utilized for assessing the consistency of S^{mid}, subsequently calculating $CS(S^{mid})$.*/
  if $FCI(S^{mid})=FCI(O,C^{mid}\cup\{d\})\neq n+1$ **then**
    $CS(S^{mid})=CS(O,C^{mid}\cup\{d\})=1$ /*$FCI(S^{mid})\neq n+1$, S^{mid} is inconsistent, so $CS(S^{mid})=1$.*/
  else
    $CS(S^{mid})=CS(O,C^{mid}\cup\{d\})=0$ /*$FCI(S^{mid})=n+1$, S^{mid} is consistent, so $CS(S^{mid})=0$.*/
  end if
  /*According to $CS(S^{mid})$, optimal scale is found by the half-search.*/
  if $CS(S^{mid})=CS(O,C^{mid}\cup\{d\})=1$ **then**
    $right=mid-1$
  else
    $left=mid+1$
  end if
end while
$OS=left-1$ /*By definition of the half-search, $left-1$ is optimal scale.*/
return OS

Table 2
Detailed time complexity of Algorithm 1.

Procedure	Function	Best complexity	Worst complexity	Average complexity
1	Calculating decision class	$O(n)$	$O(nt)$	$O(nt)$
3, 5, 7, 29	Assigning a value	$O(1)$	$O(1)$	$O(1)$
10	Calculating similarity class	$O(n)$	$O(nm)$	$O(nm)$
9-15	Calculating FCI	$O(n)$	$O(n^2m)$	$O(n^2m)$
17-21	Calculating CS	$O(1)$	$O(1)$	$O(1)$
3, 4, 5, 23-28	Half-search algorithm	$O(1)$	$O(log_2L)$	$O(log_2L)$
Total complexity	Finding optimal scale	$O(n)$	$O(n^2mlog_2L)$	$O(n^2mlog_2L)$

Table 3
Detailed time complexity of Algorithm 2.

Procedure	Function	Best complexity	Worst complexity	Average complexity
1	Calculating decision class	$O(n)$	$O(nt)$	$O(nt)$
3, 6, 32	Assigning a value	$O(1)$	$O(1)$	$O(1)$
9	Calculating similarity class	$O(n)$	$O(nm)$	$O(nm)$
8-14	Calculating FCI	$O(n)$	$O(n^2m)$	$O(n^2m)$
16-20	Calculating FCS	$O(1)$	$O(1)$	$O(1)$
5-21	Calculating OSFCS	$O(nL)$	$O(n^2mL)$	$O(n^2mL)$
23-31	Half-search algorithm	$O(1)$	$O(log_2L)$	$O(log_2L)$
Total complexity	Finding optimal scale	$O(nL)$	$O(n^2mL)$	$O(n^2mL)$

indicates that the object after FCO does not require attention and does not increase unnecessary time consumption, as portrayed in Equation (19).

$$OSFCS=[FCS(S^1),FCS(S^2),...,FCS(S^L)]=[\lambda^1,\lambda^2,...,\lambda^L],\ 0\le\lambda^1\le\lambda^2\le...\le\lambda^L\le1. \tag{18}$$

$$\lambda^1=\frac{n-FCI(S^1)+1}{L\times n},\ \lambda^{l+1}=\lambda^l+\frac{n-FCI(S^{l+1})+1}{L\times n},\ 1\le l\le L-1. \tag{19}$$

Table 4
Detailed time complexity of Algorithm 3.

Procedure	Function	Best complexity	Worst complexity	Average complexity
1	Calculating decision class	$O(n)$	$O(nt)$	$O(nt)$
3, 5, 7	Assigning a value	$O(1)$	$O(1)$	$O(1)$
10	Calculating similarity class	$O(n)$	$O(nm)$	$O(nm)$
9-15	Calculating FCI	$O(n)$	$O(n^2m)$	$O(n^2m)$
17-21	Calculating CS	$O(1)$	$O(1)$	$O(1)$
3, 4, 5, 23-28	Half-search algorithm	$O(1)$	$O(log_2(m+1))$	$O(log_2(m+1))$
Total complexity	Finding feature selection	$O(n)$	$O(n^2mlog_2(m+1))$	$O(n^2mlog_2(m+1))$

Table 5
Detailed time complexity of Algorithm 4.

Procedure	Function	Best complexity	Worst complexity	Average complexity
1	Calculating decision class	$O(n)$	$O(nt)$	$O(nt)$
3, 6	Assigning a value	$O(1)$	$O(1)$	$O(1)$
9	Calculating similarity class	$O(n)$	$O(nm)$	$O(nm)$
8-14	Calculating FCI	$O(n)$	$O(n^2m)$	$O(n^2m)$
16-20	Calculating FCS	$O(1)$	$O(1)$	$O(1)$
4-21	Calculating FSFCS	$O(nm)$	$O(n^2m^2)$	$O(n^2m^2)$
23-31	Half-search algorithm	$O(1)$	$O(log_2(m+1))$	$O(log_2(m+1))$
Total complexity	Finding feature selection	$O(nm)$	$O(n^2m^2)$	$O(n^2m^2)$

Algorithm 2: Optimal scale selection about optimal scale selection fuzzy contradictory state sequence.

Input: A $MIVDT\ S=(O,C\cup\{d\})$, $O=\{o_1,o_2,...,o_n\}$. L threshold values ς^l, L scales $l\in\{1,2,...,L\}$.
Output: Optimal scale OS and $OSFCS$.

Calculating $O/R_d=\{R_d(o)|o\in O\}=\{D_1,D_2,...,D_t\}$
/*Calculating $OSFCS$.*/
$OSFCS=[\lambda^1,\lambda^2,...,\lambda^L]$
/*Calculating $FCI(S^l)$.*/
for $l=1\rightarrow L$ **do**
  $FCI(S^l)$=n+1
  /*Verification from the first object to the final object.*/
  for $i=1\rightarrow n$ **do**
    Calculating $SC^{\varsigma^l}_{C^l}(o_i)$
    if $SC^{\varsigma^l}_{C^l}(o_i)\nsubseteq R_d(o_i)$ **then**
      $FCI(S^l)$=i
      break
    end if
  end for
  /*$FCI(S^l)$ is utilized for assessing the consistency of S^l, subsequently calculating λ^l.*/
  if $l=1$ **then**
    $\lambda^1=\frac{n-FCI(S^1)+1}{L\times n}$
  else
    $\lambda^l=\lambda^{l-1}+\frac{n-FCI(S^l)+1}{L\times n}$
  end if
end for
/*Start finding optimal scale by the half-search in $OSFCS$.*/
$left=1, right=L$
while $left\le right$ **do**
  $mid=(left+right)//2$
  if $\lambda^{mid}=0$ **then**
    $right=mid-1$
  else
    $left=mid+1$
  end if
end while
$OS=left$ /*By definition of the half-search, $left$ is optimal scale.*/
return OS and $OSFCS$

Algorithm 2 is necessary to iteratively compute the complete $OSFCS$ and identify the optimal scale by employing a half-search method.

Example 2. In Table 1, supposing $\varsigma^1=0.8, \varsigma^2=0.5$, $\zeta_1=\frac{2}{3}, \zeta_2=\frac{1}{3}$, the decision classes and similarity classes under the first scale and the second scale are acquired respectively. After calculating $O/R(d)=\{\{1,2,5,6\},\{8,3,4,7\}\}$, $FCI(S^1)=9$, $FCI(S^2)=1$, $FCS(S^1)=0$, and $FCS(S^2)=0.5$, it is calculated by Algorithm 2 that the optimal scale $OS=1$.

$$SC^{\varsigma^1}_{C^1}(o_1)=\{1\} \quad SC^{\varsigma^2}_{C^2}(o_1)=\{1,8\}$$
$$SC^{\varsigma^1}_{C^1}(o_2)=\{2\} \quad SC^{\varsigma^2}_{C^2}(o_2)=\{2,5,6\}$$
$$SC^{\varsigma^1}_{C^1}(o_3)=\{3\} \quad SC^{\varsigma^2}_{C^2}(o_3)=\{3,6,8\}$$
$$SC^{\varsigma^1}_{C^1}(o_4)=\{4\} \quad SC^{\varsigma^2}_{C^2}(o_4)=\{4,6\}$$
$$SC^{\varsigma^1}_{C^1}(o_5)=\{5\} \quad SC^{\varsigma^2}_{C^2}(o_5)=\{2,5,6\}$$
$$SC^{\varsigma^1}_{C^1}(o_6)=\{6\} \quad SC^{\varsigma^2}_{C^2}(o_6)=\{3,4,5,6\}$$
$$SC^{\varsigma^1}_{C^1}(o_7)=\{7\} \quad SC^{\varsigma^2}_{C^2}(o_7)=\{3,4,6,7,8\}$$
$$SC^{\varsigma^1}_{C^1}(o_8)=\{8\} \quad SC^{\varsigma^2}_{C^2}(o_8)=\{8\}$$

5. Feature selection for multi-scale interval valued decision table

In this section, on the basis of obtaining the optimal scale, feature selection is carried out in S^{OS}. The utilization of these algorithms in a medical context is then demonstrated.

5.1. Feature selection based on contradictory state sequence

In a $MIVDT$ $S=(O,C\cup\{d\})$, $B,H\subseteq C$, $l\in\{1,2,...,L\}$, $\varsigma^l\in[0,1]$, $C=\{c_1,c_2,...,c_m\}$. On account of *Proposition* 3, if $H\subseteq B$, it follows that $SC^{\varsigma^l}_{B^l}(o_i)\subseteq SC^{\varsigma^l}_{H^l}(o_i)$, for all $o_i\in O$. Then when $SC^{\varsigma^l}_{B^l}(o_i)\nsubseteq R_d(o_i)$, $SC^{\varsigma^l}_{H^l}(o_i)\nsubseteq R_d(o_i)$. This depicts that as the attribute set becomes a subset of it, the inconsistency does not diminish, which gratifies the essential concept of two sequences. Delete the first attribute in the conditional attribute set C successively to form a conditional attribute subset sequence, in other words, the last j attributes in C are used to constitute the conditional attribute subset, $j\in\{m,m-1,m-2,...,1,0\}$.

$$\{\{c_1,c_2,...,c_m\},\{c_2,...,c_m\},\{c_3,...,c_m\},...,\{c_m\},\varnothing\}=\{C_m,C_{m-1},C_{m-2},...,C_1,C_0\}. \tag{20}$$

The corresponding $S_j=(O,C_j\cup\{d\})$ generates $CS(S_j)$ and $FCS(S_j)$, hence making up $FSCS$ and $FSFCS$ respectively.

$$FSCS=[CS(S_m),CS(S_{m-1}),...,CS(S_1),CS(S_0)]. \tag{21}$$

In the context of $FSCS$, the positioning of 0 is consistently to the left of 1. Then the feature selection result to be sought, that is, the minimal subset C_j to ensure that the sub-information table S_j is consistent. Ultimately, the feature selection result is determined by the final sequence of 0, with the sequence commencing at 1. The obstacle of seeking the feature selection is then translated into the trouble of locating a certain number in a sequential array. When employing $FSCS$ for feature selection, the more effective half-search method is utilized.

Example 3. In Table 1, providing $\varsigma^1=\varsigma^2=0.8$, $\zeta_1=\frac{2}{3}$, $\zeta_2=\frac{1}{3}$, the decision classes and similarity classes under the second scale about C_2, C_1 and C_0 are secured respectively. After calculating $O/R(d)=\{\{1,2,5,6\},\{8,3,4,7\}\}$, $FCI(S_2)=9$, $FCI(S_1)=1$, $FCI(S_0)=1$, $CS(S_2)=0$, $CS(S_1)=1$, and $CS(S_0)=1$, as thus the feature selection result is C_2.

$$SC^{\varsigma^2}_{C_2}(o_1)=\{1\} \quad SC^{\varsigma^2}_{C_1}(o_1)=\{1,4,8\} \quad SC^{\varsigma^2}_{C_0}(o_1)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^2}_{C_2}(o_2)=\{2,5\} \quad SC^{\varsigma^2}_{C_1}(o_2)=\{2,5\} \quad SC^{\varsigma^2}_{C_0}(o_2)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^2}_{C_2}(o_3)=\{3\} \quad SC^{\varsigma^2}_{C_1}(o_3)=\{3,7\} \quad SC^{\varsigma^2}_{C_0}(o_3)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^2}_{C_2}(o_4)=\{4\} \quad SC^{\varsigma^2}_{C_1}(o_4)=\{1,4,8\} \quad SC^{\varsigma^2}_{C_0}(o_4)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^2}_{C_2}(o_5)=\{5\} \quad SC^{\varsigma^2}_{C_1}(o_5)=\{5,6\} \quad SC^{\varsigma^2}_{C_0}(o_5)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^2}_{C_2}(o_6)=\{6\} \quad SC^{\varsigma^2}_{C_1}(o_6)=\{6,7\} \quad SC^{\varsigma^2}_{C_0}(o_6)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^2}_{C_2}(o_7)=\{7\} \quad SC^{\varsigma^2}_{C_1}(o_7)=\{6,7\} \quad SC^{\varsigma^2}_{C_0}(o_7)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^2}_{C_2}(o_8)=\{8\} \quad SC^{\varsigma^2}_{C_1}(o_8)=\{1,4,8\} \quad SC^{\varsigma^2}_{C_0}(o_8)=\{1,2,3,4,5,6,7,8\}$$

5.2. Feature selection based on fuzzy contradictory state sequence

The $FSFCS$ paraphrases the transition process from consistent to inconsistent as the conditional attribute set C becomes smaller. Similarly, it is indispensable to enumerate the growth rule of FCS in $FSFCS$. The requirement for this rule is that it must show monotonicity and have a range of $[0,1]$. To boost the effectiveness of Algorithm 4, a FCS growth rule on the strength of FCI is introduced. This manifests that the object following FCO does not require attention and does not augment unnecessary time consumption, as depicted in Equation (23).

Algorithm 3: Feature selection about feature selection contradictory state sequence.

Input: A $MIVDT$ $S=(O,C\cup\{d\})$, $O=\{o_1,o_2,...,o_n\}$. OS, ς^{OS}.
Output: Feature selection C_j, $j\in\{m,m-1,m-2,...,1,0\}$.

```
Calculating O/R_d = {R_d(o)|o ∈ O} = {D_1, D_2, ..., D_t}
/*Start finding feature selection by the half-search.*/
left = 0, right = m
while left ≤ right do
    mid = (left + right)//2
    /*Calculating FCI(S_mid).*/
    FCI(S_mid) = n+1
    /*Verification from the first object to the final object.*/
    for i = 1 → n do
        Calculating SC^{ς^OS}_{C^OS_mid}(o_i)
        if SC^{ς^OS}_{C^OS_mid}(o_i) ⊈ R_d(o_i) then
            FCI(S_mid) = i
            break
        end if
    end for
    /*FCI(S_mid) is utilized for assessing the consistency of S_mid, subsequently calculating CS(S_mid).*/
    if FCI(S_mid) = FCI(O, C_mid ∪ {d}) ≠ n + 1 then
        CS(S_mid) = CS(O, C_mid ∪ {d}) = 1 /*FCI(S_mid) ≠ n + 1, S_mid is inconsistent, so CS(S_mid) = 1.*/
    else
        CS(S_mid) = CS(O, C_mid ∪ {d}) = 0 /*FCI(S_mid) = n + 1, S_mid is consistent, so CS(S_mid) = 0.*/
    end if
    /*According to CS(S_mid), feature selection is found by the half-search.*/
    if CS(S_mid) = CS(O, C_mid ∪ {d}) = 0 then
        right = mid − 1
    else
        left = mid + 1
    end if
end while
return C_{m+1−left} /*By definition of half-search, C_{m+1−left} is feature selection.*/
```

$$FSFCS=[FCS(S_m),FCS(S_{m-1}),...,FCS(S_1),FCS(S_0)]=[\lambda_m,\lambda_{m-1},...,\lambda_1,\lambda_0]. \tag{22}$$

$$\lambda_m=\frac{n-FCI(S_m)+1}{(m+1)\times n},\ \lambda_j=\lambda_{j+1}+\frac{n-FCI(S_j)+1}{(m+1)\times n},\ 0\le j\le m-1. \tag{23}$$

Algorithm 4 necessitates the sequential computation of the entire $FSFCS$ and the subsequent search for the feature selection outcome employing the half-search method.

Example 4. In Table 1, supposing $\varsigma^1=0.8$, $\varsigma^2=0.5$, $\zeta_1=\frac{2}{3}$, $\zeta_2=\frac{1}{3}$, the decision classes and similarity classes under the first scale about C_2, C_1 and C_0 are secured respectively. After calculating $O/R(d)=\{\{1,2,5,6\},\{8,3,4,7\}\}$, $FCI(S_2)=9$, $FCI(S_1)=1$, $FCI(S_0)=1$, $FCS(S_2)=0$, $FCS(S_1)=0.33$, and $FCS(S_0)=0.67$, therefore, the outcome of feature selection is C_2.

$$SC^{\varsigma^1}_{C_2}(o_1)=\{1\}\quad SC^{\varsigma^1}_{C_1}(o_1)=\{1,8\}\quad SC^{\varsigma^1}_{C_0}(o_1)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^1}_{C_2}(o_2)=\{2\}\quad SC^{\varsigma^1}_{C_1}(o_2)=\{2\}\quad SC^{\varsigma^1}_{C_0}(o_2)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^1}_{C_2}(o_3)=\{3\}\quad SC^{\varsigma^1}_{C_1}(o_3)=\{3,7\}\quad SC^{\varsigma^1}_{C_0}(o_3)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^1}_{C_2}(o_4)=\{4\}\quad SC^{\varsigma^1}_{C_1}(o_4)=\{4,8\}\quad SC^{\varsigma^1}_{C_0}(o_4)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^1}_{C_2}(o_5)=\{5\}\quad SC^{\varsigma^1}_{C_1}(o_5)=\{5,6\}\quad SC^{\varsigma^1}_{C_0}(o_5)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^1}_{C_2}(o_6)=\{6\}\quad SC^{\varsigma^1}_{C_1}(o_6)=\{6\}\quad SC^{\varsigma^1}_{C_0}(o_6)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^1}_{C_2}(o_7)=\{7\}\quad SC^{\varsigma^1}_{C_1}(o_7)=\{6,7\}\quad SC^{\varsigma^1}_{C_0}(o_7)=\{1,2,3,4,5,6,7,8\}$$
$$SC^{\varsigma^1}_{C_2}(o_8)=\{8\}\quad SC^{\varsigma^1}_{C_1}(o_8)=\{1,8\}\quad SC^{\varsigma^1}_{C_0}(o_8)=\{1,2,3,4,5,6,7,8\}$$

The process of feature selection for $MIVDT$ is outlined, followed by a specific application scenario. In this context, healthcare organizations utilize diagnostic data to forecast patients' disease risk, incorporating attributes such as age, gender, blood pressure, and blood sugar levels. Both the blood pressure and blood sugar levels recorded in the dataset are represented as interval values. The original dataset may contain missing values or outliers, necessitating preprocessing. To streamline the decision table and eliminate redundant and irrelevant attributes, a feature selection algorithm is employed. Subsequently, the simplified multi-scale interval value decision table serves as the training input for the disease risk prediction model. Upon assessing the algorithm's efficiency and accuracy, disease risk prediction for new patients based on medical records can be conducted.

Algorithm 4: Feature selection about feature selection fuzzy contradictory state sequence.

Input: A $MIVDT\ S=(O,C\cup\{d\})$, $O=\{o_1,o_2,...,o_n\}$. OS, ς^{OS}.
Output: Feature selection C_j, $j\in\{m,m-1,m-2,...,1,0\}$.

Calculating $O/R_d=\{R_d(o)|o\in O\}=\{D_1,D_2,...,D_t\}$
/*Calculating $FSFCS$.*/
$FSFCS=[\lambda_m,\lambda_{m-1},...,\lambda_1,\lambda_0]$
for *all* $j\in\{m,m-1,...,1,0\}$ **do**
  /*Calculating $FCI(S_j)$.*/
  $FCI(S_j)$=n+1
  /*Verification from the first object to the final object.*/
  for $i=1\rightarrow n$ **do**
    Calculating $SC^{\varsigma^{OS}}_{C^{OS}_j}(o_i)$
    if $SC^{\varsigma^{OS}}_{C^{OS}_j}(o_i)\nsubseteq R_d(o_i)$ **then**
      $FCI(S_j)$=i
      break
    end if
  end for
  /*$FCI(S_j)$ is utilized for assessing the consistency of S_j, subsequently calculating λ_j.*/
  if $j=m$ **then**
    $\lambda_m=\frac{n-FCI(S_m)+1}{(m+1)\times n}$
  else
    $\lambda_j=\lambda_{j+1}+\frac{n-FCI(S_j)+1}{(m+1)\times n}$
  end if
end for
/*Start finding feature selection by the half-search in $FSFCS$.*/
$left,right=0,m$
while $left\le right$ **do**
  $mid=(left+right)//2$
  if $\lambda_{mid}=0$ **then**
    $right=mid-1$
  else
    $left=mid+1$
  end if
end while
return $C_{m+1-left}$ /*By definition of half-search, $C_{m+1-left}$ is feature selection.*/

Table 6
Data sets information.

Datasets	Abbreviations	Objects	Features	Classes
CervicalCancerBehaviorRisk	CCBR	72	19	2
DiferentiatedThyroidCancerRecurrence	DTCR	383	16	2
WholeScaleCustomers	WSC	440	7	3
NationalPollonHealthyAging(NPHA)	NPHA	714	14	3
WirelessIndoorLocalization	WIL	2000	7	4
AuctionVerifcation	AV	2043	7	2
EstimationofObesityLevelsBasedon-EatingHabitsandPhysicalCondition	EOLBEHPC	2111	16	7
PredictStudents'DropoutandAcademicSuccess	PSDAS	4424	36	3
Tunadromd	Tunadromd	4465	**242**	2
TarvelReviewRatings	TRR	5456	25	**9**
NationalHealthandNutritionHealth-Survey2013-2014AgePredictionSubset	NHNHSAPS	6287	7	2
ShillBiddingDataset	SBD	6321	13	2
ElectricalGridStabilitySimulated	EGSS	10000	13	2
A1412020PredictiveMaintenance	APM	10000	14	3
NATICUSdroid(AndroidPermissions)Dataset	NATICUSdroid	29333	86	2
AccelerometerGyroMobilePhoneDataset	AGMPD	31991	8	2
SepsisSurvivalMinimalClinicalRecords	SSMCR	**110341**	3	2

6. Experimental analysis

In this section, the feasibility and effectiveness of the four algorithms are validated. All experimental hardware setups are configured as Windows 11, Intel(R) Core(TM) i7-10750H CPU @ 2.60 GHz, and 16.0 GB memory. The software environment for executing the algorithm is Python 3.7. Seventeen open data sets from the University of California, Irvine were selected to verify the algorithms presented in this article. Details of these seventeen open data sets are shown in Table 6, later, the data set names will be represented by their abbreviations. The method in [11] is used to convert ordinary data sets to MIS, which is then uniformly processed to $MIVDT$. The detailed steps are as follows.

Table 7
Optimal scale correlation experiments with divergent data sets about the first parameter combination.

Data sets	PT (s)	TSS	Time1 (s)	OS	Time2 (s)	OSFCS
DTCR	**1.52**	306	0.23	1	0.26	[0.00, 0.13, 0.33, 0.53, 0.73]
NPHA	4.52	496	0.07	1	0.09	[0.00, 0.20, 0.40, 0.60, 0.80]
AV	27.00	1330	0.10	1	0.14	[0.00, 0.20, 0.40, 0.60, 0.80]
EOLBEHPC	35.05	1677	0.28	1	0.33	[0.00, 0.20, 0.40, 0.60, 0.80]
PSDAS	107.36	3538	1.81	1	2.20	[0.00, 0.19, 0.39, 0.59, 0.79]
Tunadromd	813.52	3510	5.99	1	7.97	[0.00, 0.20, 0.40, 0.60, 0.80]
TRR	194.75	4345	2.20	1	2.46	[0.00, 0.20, 0.40, 0.60, 0.80]
NHNHSAPS	35.40	1697	0.14	1	0.20	[0.00, 0.20, 0.40, 0.60, 0.80]
SBD	180.20	5041	0.35	1	0.52	[0.00, 0.20, 0.40, 0.60, 0.80]
NATICUSdroid	22317.00	19191	13.86	1	18.41	[0.00, 0.20, 0.40, 0.60, 0.80]
AGMPD	6068.10	9716	1909.24	**5**	8731.77	**[0.00, 0.00, 0.00, 0.00, 0.00]**
SSMCR	44597.90	**221**	**0.01**	1	**0.01**	[0.00, 0.20, 0.40, 0.60, 0.80]

Table 8
Feature selection correlation experiments with divergent data sets about the first parameter combination.

Data sets	Time3 (s)	AR	ARR	Time4 (s)	FSFCS	Accuracy (%)
DTCR	4.28	C_{10}	0.63	10.56	*	**99.98**
NPHA	0.06	C_{13}	0.93	0.14	*	99.01
AV	0.10	C_{7}	1.00	0.13	*	99.85
EOLBEHPC	1.37	C_{16}	1.00	1.86	*	**99.99**
PSDAS	176.19	C_{36}	1.00	381.90	*	**99.99**
Tunadromd	463.49	C_{55}	0.23	34567.71	*	98.79
TRR	19.97	C_{24}	0.96	41.64	*	**99.98**
NHNHSAPS	0.04	C_{7}	1.00	0.09	*	99.12
SBD	0.21	C_{9}	0.69	0.38	*	99.92
NATICUSdroid	660.86	C_{86}	1.00	1655.24	*	99.68
AGMPD	38.81	C_{0}	**0.00**	6068.78	**[0, 0, 0, 0, 0, 0, 0]**	98.22
SSMCR	**0.01**	C_{3}	1.00	**0.01**	[0.00, 0.25, 0.50, 0.75]	**99.98**

- Initially, the seventeen datasets must undergo preprocessing, which entails the conversion of character data into numerical data, standardization of missing values, and elimination of irrelevant attribute columns such as user ID. The absent data point is substituted with the mean value of the dataset within the respective column. Identical character data is equated with sole numeric data, with the numeric values commencing from zero.
- Expanding the scale infinitely will lead to a reduction in the distance between adjacent scales, resulting in a convergence of roughness. Therefore transforming single-scale data into five-scale interval value data for each conditional attribute by the method of [11]. As well as the minimum value of each column is the lower bound of the interval value in this column, plus, the elements are the upper bound of the interval value.
- For unclassified datasets, k-means clustering is required to group them into nine classes while limiting the maximum number of iterations to ten.

6.1. Evaluating the performance of four algorithms

All experimental results are derived from average values obtained using the five-fold cross-validation method. This approach not only mitigates the inherent variability in evaluating model performance, yielding more dependable performance estimates, but also bolsters the generalizability and stability of the model.

The algorithm running process involves multiple parameters, and different combinations of these parameters are used to compare the performance of the algorithm. The first parameter combination is $(\zeta_1 = 1/2, \zeta_2 = 1/2, \varsigma^1 = 0.9, \varsigma^2 = 0.7, \varsigma^3 = 0.5, \varsigma^4 = 0.3, \varsigma^5 = 0.1)$, while the second parameter combination is $(\zeta_1 = 2/3, \zeta_2 = 1/3, \varsigma^1 = 0.8, \varsigma^2 = 0.8, \varsigma^3 = 0.8, \varsigma^4 = 0.8, \varsigma^5 = 0.8)$. The first and second parameter combinations are applied to four algorithms and the outcomes are presented in Table 7 ∼ Table 10. (FC-The algorithm under the first parameter combination, SC-The algorithm under the second parameter combination, PT-Consistent processing time, TSS-Training set size after consistency processing, Time1-Finding optimal scale time, OS-Optimal scale, Time2-Find OSFCS time, Time3-Feature selection time, AR-Feature selection, ARR-Feature selection rate, Time4-Finding FSFCS time.)

The symbol $*$ in the table denotes information that exceeds the available space. Through comparison, it can be observed that the algorithm under the first parameter combination exhibits better time consumption, accuracy and attribute reduction rate than the algorithm under the second parameter combination in most cases. The time consumption comparison diagram under two parameter combinations are exhibited in Fig. 5 and Fig. 6, and the accuracy comparison diagram is displayed in Fig. 7. It is important to acknowledge that optimal parameter combinations may vary across different application contexts and therefore require further investigation during practical implementation.

Table 9
Optimal scale correlation experiments with divergent data sets about the second parameter combination.

Data sets	PT (s)	TSS	Time1 (s)	OS	Time2 (s)	OSFCS
DTCR	**1.65**	302	0.73	1	0.76	[0.00, 0.19, 0.39, 0.59, 0.79]
NPHA	5.10	187	0.02	1	0.04	[0.00, 0.20, 0.40, 0.60, 0.80]
AV	33.07	537	0.03	1	0.05	[0.00, 0.20, 0.40, 0.60, 0.80]
EOLBEHPC	38.22	1448	0.31	1	0.35	[0.00, 0.20, 0.40, 0.60, 0.80]
PSDAS	145.23	3520	1.04	1	1.40	[0.00, 0.20, 0.40, 0.60, 0.80]
Tunadromd	961.58	3510	6.10	1	8.18	[0.00, 0.20, 0.40, 0.60, 0.80]
TRR	262.76	4271	2.80	1	3.14	[0.00, 0.20, 0.40, 0.60, 0.80]
NHNHSAPS	49.98	1112	0.10	1	0.14	[0.00, 0.20, 0.40, 0.60, 0.80]
SBD	238.11	4896	0.39	1	0.57	[0.00, 0.20, 0.40, 0.60, 0.80]
NATICUSdroid	23263.77	19191	14.59	1	19.66	[0.00, 0.20, 0.40, 0.60, 0.80]
AGMPD	11569.01	5447	1265.79	**5**	6074.15	**[0.00, 0.00, 0.00, 0.00, 0.00]**
SSMCR	45273.66	**22**	**0.01**	1	**0.01**	[0.00, 0.20, 0.40, 0.60, 0.80]

Table 10
Feature selection correlation experiments with divergent data sets about the second parameter combination.

Data sets	Time3 (s)	AR	ARR	Time4 (s)	FSFCS	Accuracy (%)
DTCR	1.31	C_{16}	1.00	3.24	*	99.48
NPHA	0.02	C_{13}	0.93	0.04	*	93.50
AV	0.03	C_{7}	1.00	0.04	*	95.00
EOLBEHPC	0.60	C_{16}	1.00	1.10	*	95.45
PSDAS	7.70	C_{36}	1.00	15.06	*	99.94
Tunadromd	445.05	C_{55}	0.23	40625.11	*	98.79
TRR	5.71	C_{24}	0.96	15.59	*	99.87
NHNHSAPS	0.12	C_{7}	1.00	0.07	*	86.43
SBD	0.12	C_{9}	0.69	0.29	*	98.60
NATICUSdroid	664.33	C_{86}	1.00	1620.49	*	99.68
AGMPD	31.34	C_{0}	**0.00**	4933.21	**[0, 0, 0, 0, 0, 0, 0]**	98.22
SSMCR	**0.01**	C_{3}	1.00	**0.01**	[0.00, 0.25, 0.50, 0.75]	**99.98**

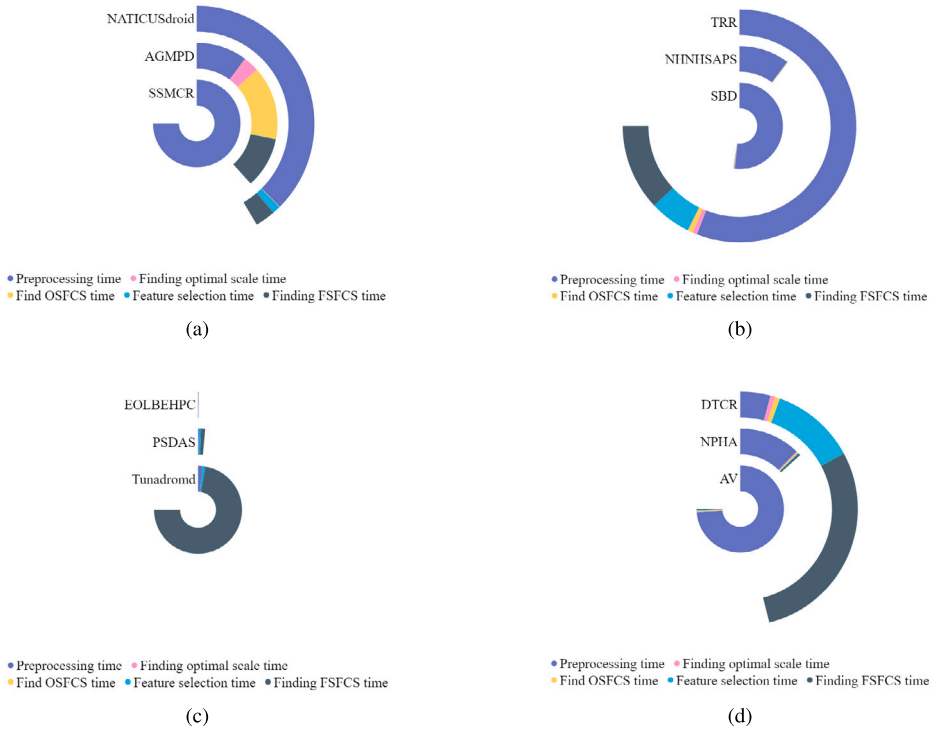

Fig. 5. The time consumption (s) of the first parameter combination ($\zeta_1 = \zeta_2 = 1/2, \varsigma^1 = 0.9, \varsigma^2 = 0.7, \varsigma^3 = 0.5, \varsigma^4 = 0.3, \varsigma^5 = 0.1$).

Fig. 6. The time consumption (s) of the second parameter combination ($\zeta_1 = 2/3, \zeta_2 = 1/3, \varsigma^1 = \varsigma^2 = \varsigma^3 = \varsigma^4 = \varsigma^5 = 0.8$).

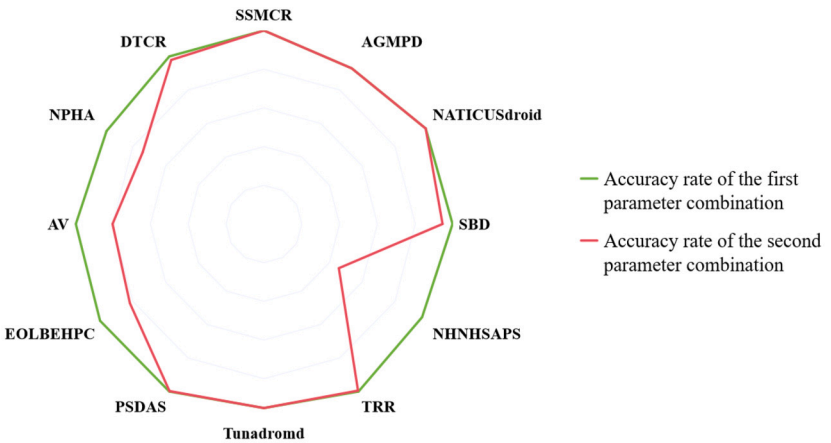

Fig. 7. Accuracy (%) of the algorithms in this paper under the two parameter combinations.

6.2. Contrast test

In this subsection, five commonly used machine learning algorithms (see Fig. 8), namely Logistic Regression Model(LR), Decision Tree Model($TREE$), Support Vector Machine Model(SVM), Random Forest Model($FOREST$) and Gradient Boosted Decision Tree Model($GBDT$) are compared with the algorithms of this paper. The threshold value was set at a level of 1×10^{-7}, the number of base evaluators was set as 100, and all other parameters were initialized with default values. Accuracy, precision, recall, F1, AUC and other evaluation indexes were recorded for binary classification datasets; whereas only accuracy was recorded for non-binary classification datasets. Table 11 ∼ Table 15 present corresponding results across five machine learning algorithms. The appearance of "null" in the table signifies either prolonged processing time or unattainability of outcomes. In its entirety, the algorithm outlined in this paper demonstrates a level of superiority over the aforementioned machine learning algorithms.

In addition to the above machine learning related algorithms, several rough set algorithms are also compared, including the interval-valued based weighted neighborhood rough set ($IVWNRS$) [37], interval-valued based neighborhood rough set ($IVNRS$), local neighborhood rough set ($LCER$) [24], distance measure based fuzzy rough set ($AVDP$) [23] ($\gamma = 0.005$), multi-granularity attribute selector ($MGAS$) [17], bucket and attribute group based neighborhood rough set ($BAGR$) [4] and multi-level

Table 11
LR corresponding metrics for different datasets.

Data sets	Objects	Accuracy (%)	Precision (%)	Recall rate (%)	F1 score (%)	AUC (%)
DTCR	383	85.62	82.11	77.18	74.34	98.32
NPHA	714	51.82	null	null	null	null
AV	2043	86.05	37.18	5.71	8.94	74.50
EOLBEHPC	2111	86.89	null	null	null	null
PSDAS	4424	76.74	null	null	null	null
Tunadromd	4465	98.45	98.81	99.27	99.04	99.75
TRR	5456	96.86	null	null	null	null
NHNHSAPS	6287	**99.43**	**99.38**	99.95	**99.66**	**99.97**
SBD	6321	97.72	88.15	90.96	89.48	99.60
NATICUSdroid	29333	95.40	95.54	95.26	95.39	98.65
AGMPD	31991	98.06	98.24	99.81	99.02	86.78
SSMCR	110341	92.65	92.65	**100.00**	96.18	70.68

Table 12
TREE corresponding metrics for different datasets.

Data sets	Objects	Accuracy (%)	Precision (%)	Recall rate (%)	F1 score (%)	AUC (%)
DTCR	383	81.63	78.66	88.10	79.34	83.68
NPHA	714	40.75	null	null	null	null
AV	2043	76.33	67.46	93.53	70.25	83.67
EOLBEHPC	2111	92.47	null	null	null	null
PSDAS	4424	68.11	null	null	null	null
Tunadromd	4465	98.79	98.82	99.69	99.25	98.06
TRR	5456	83.59	null	null	null	null
NHNHSAPS	6287	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
SBD	6321	99.68	98.96	98.07	98.50	98.98
NATICUSdroid	29333	93.90	96.72	90.90	93.59	94.51
AGMPD	31991	96.16	98.50	97.58	98.01	57.81
SSMCR	110341	92.61	92.65	**100.00**	96.16	69.23

Table 13
SVM corresponding metrics for different datasets.

Data sets	Objects	Accuracy (%)	Precision (%)	Recall rate (%)	F1 score (%)	AUC (%)
DTCR	383	84.84	79.99	76.32	73.52	93.49
NPHA	714	52.24	null	null	null	null
AV	2043	79.06	30.97	20.99	18.18	59.64
EOLBEHPC	2111	87.51	null	null	null	null
PSDAS	4424	75.97	null	null	null	null
Tunadromd	4465	98.54	**99.47**	98.71	**99.08**	**99.80**
TRR	5456	94.77	null	null	null	null
NHNHSAPS	6287	97.41	97.79	99.16	98.47	99.60
SBD	6321	**98.61**	91.56	95.85	93.64	**99.80**
NATICUSdroid	29333	93.86	96.48	91.04	93.46	98.24
AGMPD	31991	98.17	98.21	**99.96**	**99.08**	82.65
SSMCR	110341	null	null	null	null	null

Table 14
FOREST corresponding metrics for different datasets.

Data sets	Objects	Accuracy (%)	Precision (%)	Recall rate (%)	F1 score (%)	AUC (%)
DTCR	383	89.26	87.27	84.46	83.23	99.04
NPHA	714	43.42	null	null	null	null
AV	2043	83.86	69.21	79.43	68.35	92.56
EOLBEHPC	2111	93.76	null	null	null	null
PSDAS	4424	77.46	null	null	null	null
Tunadromd	4465	99.48	99.67	99.69	99.68	98.75
TRR	5456	90.34	null	null	null	null
NHNHSAPS	6287	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
SBD	6321	99.34	95.46	99.51	96.95	99.56
NATICUSdroid	29333	94.78	97.39	92.05	94.49	98.77
AGMPD	31991	98.13	98.26	98.82	98.52	89.01
SSMCR	110341	null	null	null	null	null

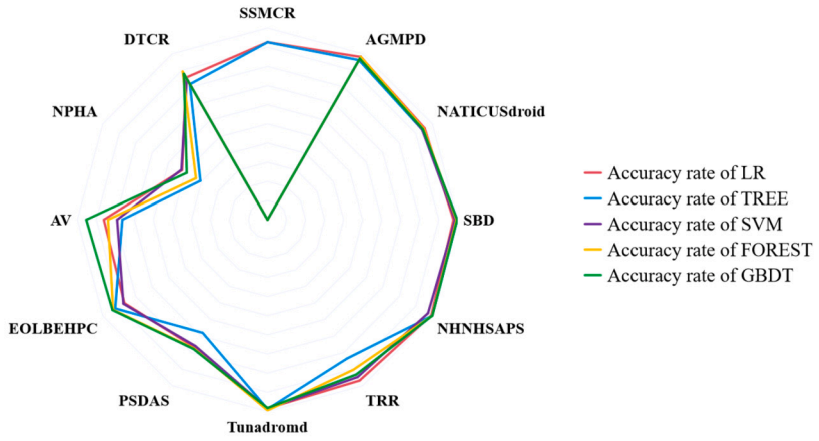

Fig. 8. Accuracy (%) of five machine learning algorithms.

Table 15
GBDT corresponding metrics for different datasets.

Data sets	Objects	Accuracy (%)	Precision (%)	Recall rate (%)	F1 score (%)	AUC (%)
DTCR	383	87.93	84.51	89.00	83.97	97.91
NPHA	714	49.01	null	null	null	null
AV	2043	95.30	82.91	81.38	81.95	94.75
EOLBEHPC	2111	94.13	null	null	null	null
PSDAS	4424	77.80	null	null	null	null
Tunadromd	4465	98.23	98.62	99.19	98.90	99.85
TRR	5456	93.25	null	null	null	null
NHNHSAPS	6287	**100.00**	**100.00**	**100.00**	**100.00**	**100.00**
SBD	6321	99.72	98.68	98.67	98.67	99.98
NATICUSdroid	29333	94.29	95.49	92.98	94.11	98.52
AGMPD	31991	97.01	98.26	98.71	98.47	91.28
SSMCR	110341	null	null	null	null	null

Table 16
Comparison of the classification accuracy (%) of different algorithms.

Data set	LR	TREE	SVM	FOREST	GBDT	FC	SC
DTCR	85.62	81.63	84.84	89.26	87.93	**99.98**	99.48
NPHA	51.82	40.75	52.24	43.42	49.01	**99.01**	93.50
AV	86.05	76.33	79.06	83.86	95.30	**99.85**	95.00
EOLBEHPC	86.89	92.47	87.51	93.76	94.13	**99.99**	95.45
PSDAS	76.74	68.11	75.97	77.46	77.80	**99.99**	99.94
Tunadromd	98.45	98.79	98.54	**99.48**	98.23	98.79	98.79
TRR	96.86	83.59	94.77	90.34	93.25	**99.98**	99.87
NHNHSAPS	99.43	**100.00**	97.41	**100.00**	**100.00**	99.12	86.43
SBD	97.72	99.68	98.61	99.34	99.72	**99.92**	98.60
NATICUSdroid	95.40	93.90	93.86	94.78	94.29	**99.68**	99.68
AGMPD	98.06	96.16	98.17	98.13	97.01	**98.22**	98.22
SSMCR	92.65	92.61	null	null	null	**99.98**	**99.98**

neighborhood based sequential three-way decision ($MNS3WD$) [33]. The relevant comparison results are presented in Table 17 and plotted in Fig. 10. The illustration demonstrates that the algorithm proposed in this paper exhibits superior performance compared to alternative algorithms. Subsequent efforts will involve conducting more comprehensive mathematical evaluations to substantiate the algorithm superiority.

6.3. Statistical analysis

The experimental findings presented above are depicted in Table 16 and Fig. 9, providing a comparison of accu- racy through a tabular and graphical representation. In this subsection, the statistical performances of seventeen datasets are systematically explored in terms of classification accuracy under divergent models, and Friedman test is carried out. The Friedman statistic is illustrated as follows.

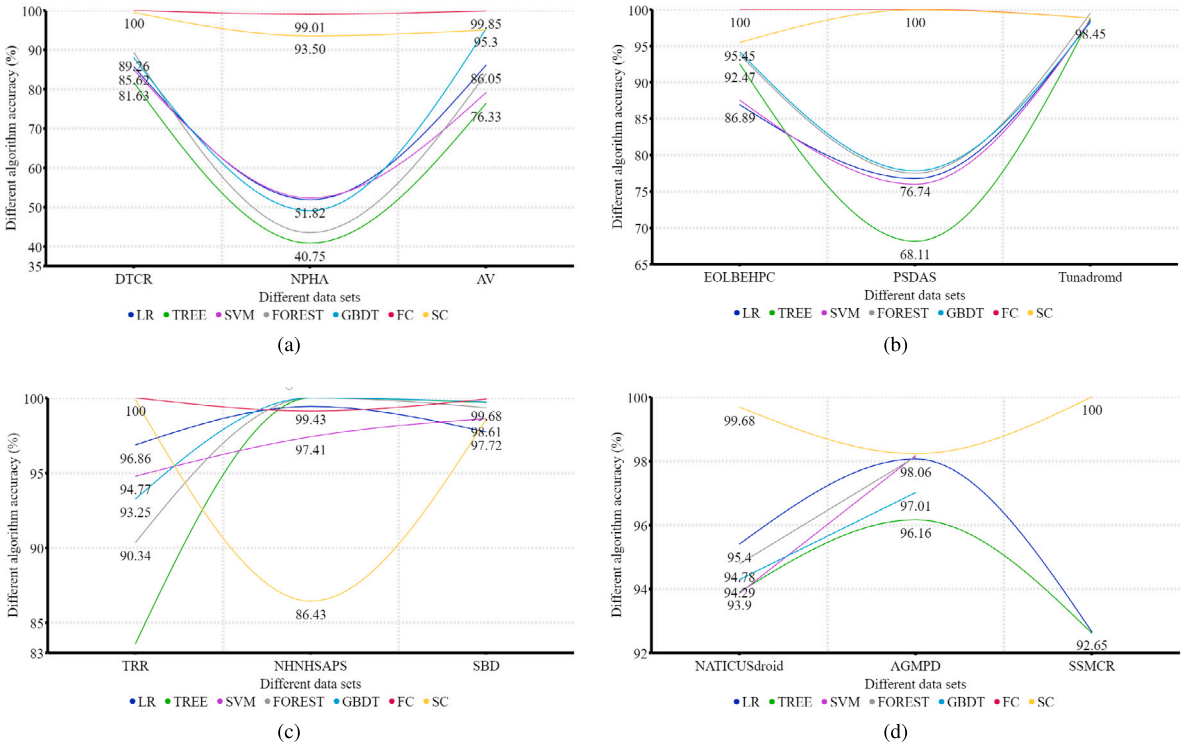

Fig. 9. Accuracy (%) of different algorithms in twelve datasets.

Table 17
Classification accuracy (%) of different algorithms.

Datasets	FC	IVWNRS	IVNRS	LCER	AVDP	MGAS	BAGR	MNS3WD
CCBR	**98.67**	91.60	89.10	76.60	62.00	82.30	82.10	77.30
WSC	**91.59**	60.20	59.80	60.00	65.50	60.00	60.00	65.50
WIL	**99.10**	97.60	97.60	97.90	55.10	**98.70**	98.40	75.00
EOLBEHPC	**99.99**	92.40	92.40	78.30	13.60	90.50	82.00	17.20
EGSS	**99.98**	**100.00**	**100.00**	95.20	54.90	90.10	95.10	60.60
APM	**98.91**	48.50	48.50	50.20	45.50	50.20	50.20	46.40

Table 18
Machine learning algorithms rank mean.

Algorithm	Rank mean
FC	6.38
SC	5.21
LR	3.33
TREE	2.58
SVM	2.75
FOREST	3.92
GBDT	3.83

$$\chi_F^2 = \frac{12N}{k(k+1)}\left(\sum_{j=1}^{k} R_j^2 - \frac{k(k+1)^2}{4}\right), \; F_F = \frac{(N-1)\,\chi_F^2}{N(k-1)-\chi_F^2}. \tag{24}$$

Where N delegates the number of data sets, while k reveals the number of methods, R indicates the average ranking of a certain approach and F manifests a F-distribution with $(k-1)$ and $(k-1)(N-1)$ degrees of freedom. Then the critical difference is demonstrated as follows. Here α suggests the significance level and q_α denotes a crucial value, the Friedman tests of machine learning algorithms and rough set algorithms are manifested in Table 18 ~ Table 21.

$$CD_\alpha = q_\alpha \sqrt{\frac{k(k+1)}{6N}}. \tag{25}$$

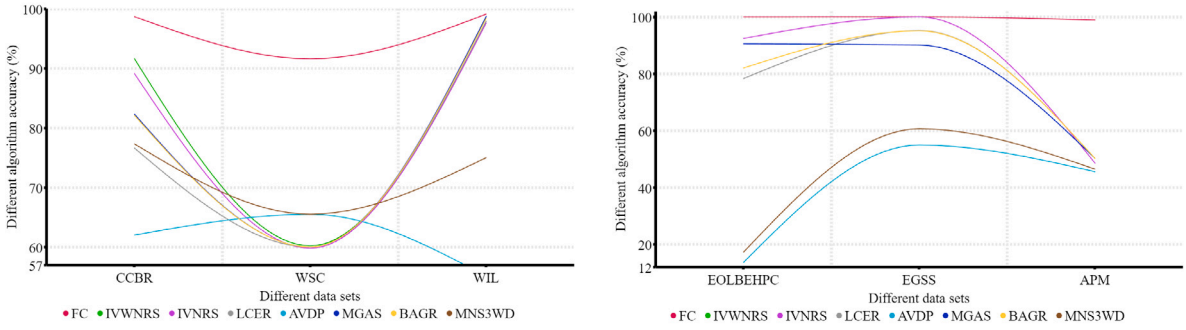

Fig. 10. Accuracy (%) of different algorithms in six datasets.

Table 19
Rough set algorithms rank mean.

Algorithm	Rank mean
FC	7.83
IVWNRS	5.42
IVNRS	4.58
LCER	4.00
AVDP	1.92
MGAS	4.83
BAGR	4.50
MNS3WD	2.92

Table 20
Machine learning algorithms Friedman test.

Friedman statistical test	Result
Number of cases	12
χ^2	29.32
Degree of freedom	6
Asymptotic significance	0.00

Table 21
Rough set algorithms Friedman test.

Friedman statistical test	Result
Number of cases	6
χ^2	22.21
Degree of freedom	7
Asymptotic significance	0.00

The null hypothesis of Friedman test posits that accuracy of the models originates from an identical population distribution. The depicted tabulations represent the test outcomes, with $\chi^2 = 29.32,\ P < 0.01$, and $\chi^2 = 22.21,\ P < 0.01$, leading to rejection of null hypothesis. This indicates a statistically momentous deviation in accuracy among the models. According to the results of Friedman statistical test, the efficiency of the algorithms provided in this paper is indeed better than other methods.

7. Conclusions

This study introduces novel approaches for feature selection in $MIVDT$, utilizing CSS and $FCSS$. The incorporation of these concepts enables a more precise and efficient description of the consistency of $MIVDT$, leading to improved effectiveness in feature selection processes. The results of the experiment demonstrate that the algorithm under consideration is capable of efficiently identifying novel feature subsets in a shorter duration and minimizing storage requirements, all while maintaining satisfactory classification accuracy. In summary, the study discussed in this paper effectively addresses the challenges encountered by conventional approaches, demonstrating superior computational efficiency and accuracy. It offers a significant point of reference for further research and practical applications in relevant domains. Subsequent studies may delve into the utilization of CS and FCS in addressing additional intricate decision-making scenarios, as well as investigate the integration of the suggested approach with other sophisticated feature selection methodologies to enhance the model's performance and generalizability. Furthermore, it proves challenging to analyze

the optimal parameter combination's value range across various scenarios. Moreover, real-world application data often exhibits increased complexity and diversity, the effective implementation strategy of the proposed method in the actual environment needs to be further explored.

CRediT authorship contribution statement

Xiaoyan Zhang: Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Zihan Feng:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation.

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

No data was used for the research described in the article.

Acknowledgement

The authors would like to thank the Associate Editor and the reviewers for their insightful comments and suggestions.

This work was supported by the National Natural Science Foundation of China (Grant NO.12371465, 62376229) and the Chongqing Natural Science Foundation of China (Grant NO.CSTB2023NSCO-MSX1063).

References

[1] M.C. Barbieri, B.I. Grisci, M. Dorn, Analysis and comparison of feature selection methods towards performance and stability, Expert Syst. Appl. (2024) 123667.
[2] B. Chen, X. Zhang, Z. Yuan, Two-dimensional improved attribute reductions based on distance granulation and condition entropy in incomplete interval-valued decision systems, Inf. Sci. 657 (2024) 119910.
[3] X. Chen, M. Luo, Incremental information fusion in the presence of object variations for incomplete interval-valued data based on information entropy, Inf. Sci. 667 (2024) 120479.
[4] Y. Chen, K. Liu, J. Song, et al., Attribute group for attribute reduction, Inf. Sci. 535 (2020) 64–80.
[5] Y. Ding, W. Xu, W. Ding, et al., IFCRL: interval-intent fuzzy concept re-cognition learning model, IEEE Trans. Fuzzy Syst. 32 (2024) 3581–3593.
[6] A. Gudyś, M. Sikora, Ł. Wróbel, Separate and conquer heuristic allows robust mining of contrast sets in classification, regression, and survival data, Expert Syst. Appl. (2024) 123376.
[7] Y. Gui, W. Gan, Y. Wu, et al., Privacy Preserving Rare Itemset Mining, Information Sciences, 2024, p. 120262.
[8] D. Guo, W. Xu, W. Ding, et al., Concept-cognitive learning survey: mining and fusing knowledge from data, Inf. Fusion (2024) 102426.
[9] D. Guo, W. Xu, Y. Qian, et al., M-FCCL: memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion, Inf. Fusion 100 (2023) 101962.
[10] J. He, L. Qu, P. Wang, et al., An oscillatory particle swarm optimization feature selection algorithm for hybrid data based on mutual information entropy, Appl. Soft Comput. 152 (2024) 111261.
[11] Z. Huang, J. Li, W. Dai, et al., Generalized multi-scale decision tables with multi-scale decision attributes, Int. J. Approx. Reason. 115 (2019) 194–208.
[12] R.M. Hussien, A.A. Abohany, A.A. Abd El-Mageed, et al., Improved binary meerkat optimization algorithm for efficient feature selection of supervised learning classification, Knowl.-Based Syst. 292 (2024) 111616.
[13] A. Kiersztyn, D. Pylak, M. Horodelski, et al., Random Clustering-Based Outlier Detector, Information Sciences, 2024, p. 120498.
[14] Q. Kong, W. Wang, W. Xu, et al., A method of data analysis based on division-mining-fusion strategy, Inf. Sci. 666 (2024) 120450.
[15] S. Li, Q. Chen, Z. Liu, et al., Bi-SGTAR: a simple yet efficient model for circRNA-disease association prediction based on known association pair only, Knowl.-Based Syst. 291 (2024) 111622.
[16] C. Liu, S. Yang, A two-stage clustering ensemble algorithm applicable to risk assessment of railway signaling faults, Expert Syst. Appl. (2024) 123500.
[17] K. Liu, X. Yang, H. Fujita, et al., An efficient selector for multi-granularity attribute reduction, Inf. Sci. 505 (2019) 457–472.
[18] W. Li, H. Zhou, W. Xu, et al., Interval dominance-based feature selection for interval-valued ordered data, IEEE Trans. Neural Netw. Learn. Syst. 34 (2023) 6898–6912.
[19] Y. Pan, W. Xu, Q. Ran, An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets, Int. J. Mach. Learn. Cybern. 14 (4) (2023) 1217–1233.
[20] D. Paulraj, K.A.M. Junaid, T. Sethukarasi, et al., A novel efficient rank-revealing QR matrix and Schur decomposition method for big data mining and clustering (RRQR-SDM), Inf. Sci. 657 (2024) 119957.
[21] C. Rao, X. Wei, X. Xiao, et al., Oversampling method via adaptive double weights and Gaussian kernel function for the transformation of unbalanced data in risk assessment of cardiovascular disease, Inf. Sci. (2024) 120410.
[22] L. Sun, Y. Ma, W. Ding, et al., LSFSR: local label correlation-based sparse multilabel feature selection with feature redundancy, Inf. Sci. (2024) 120501.
[23] C. Wang, Y. Huang, M. Shao, et al., Fuzzy rough set-based attribute reduction using distance measures, Knowl.-Based Syst. 164 (2019) 205–212.
[24] Q. Wang, Y. Qian, X. Liang, et al., Local neighborhood rough set, Knowl.-Based Syst. 153 (2018) 53–64.
[25] J. Wan, H. Chen, T. Li, et al., Interactive and complementary feature selection via fuzzy multigranularity uncertainty measures, IEEE Trans. Cybern. 53 (2) (2021) 1208–1221.
[26] W.Z. Wu, Y. Leung, Optimal scale selection for multi-scale decision tables, Int. J. Approx. Reason. 54 (8) (2013) 1107–1129.
[27] Z.H. Xie, W.Z. Wu, L.X. Wang, et al., Entropy based optimal scale selection and attribute reduction in multi-scale interval-set decision tables, Int. J. Mach. Learn. Cybern. (2024) 1–22.
[28] D. Xu, H. Peng, Y. Tang, et al., Hierarchical spatio-temporal graph convolutional neural networks for traffic data imputation, Inf. Fusion (2024) 102292.
[29] K. Xu, T. Li, M.S. Khan, et al., Body composition assessment with limited field-of-view computed tomography: a semantic image extension perspective, Med. Image Anal. 88 (2023) 102852.

[30] W. Xu, K. Cai, D.D. Wang, A novel information fusion method using improved entropy measure in multi-source incomplete interval-valued datasets, Int. J. Approx. Reason. 164 (2024) 109081.
[31] W. Xu, D. Guo, J. Mi, et al., Two-way concept-cognitive learning via concept movement viewpoint, IEEE Trans. Neural Netw. Learn. Syst. 34 (2023) 6798–6812.
[32] W. Xu, M. Huang, Z. Jiang, et al., Graph-based unsupervised feature selection for interval-valued information system, IEEE Trans. Neural Netw. Learn. Syst. (2023).
[33] X. Yang, T. Li, D. Liu, et al., A multilevel neighborhood sequential decision approach of three-way granular computing, Inf. Sci. 538 (2020) 119–141.
[34] Y. Yang, D. Chen, Z. Ji, et al., A two-way accelerator for feature selection using a monotonic fuzzy conditional entropy, Fuzzy Sets Syst. (2024) 108916.
[35] R. Zhang, Z. Meng, H. Wang, et al., Hyperscale data analysis oriented optimization mechanisms for higher education management systems platforms with evolutionary intelligence, Appl. Soft Comput. 155 (2024) 111460.
[36] X. Zhang, D. Guo, W. Xu, Two-way concept-cognitive learning with multi-source fuzzy context, Cogn. Comput. 15 (5) (2023) 1526–1548.
[37] X. Zhang, Z. Jiang, W. Xu, Feature selection using a weighted method in interval-valued decision information systems, Appl. Intell. 53 (9) (2023) 9858–9877.
[38] X. Zhang, J. Li, Incremental feature selection approach to interval-valued fuzzy decision information systems based on λ-fuzzy similarity self-information, Inf. Sci. 625 (2023) 593–619.