Engineering Applications of Artificial Intelligence 142 (2025) 109935

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

Research paper

Preference ranking organization method for enrichment evaluation-based feature selection for multiple source ordered information systems

Weihua Xu *, Zishuo Yang

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

ARTICLE INFO

Keywords:
Multi-source
Feature selection
Ordered decision information system
Rough set

ABSTRACT

Feature selection in multi-source ordered decision-making information systems is an underexplored domain that poses unique challenges. This paper introduces a novel approach to fill this research gap by integrating the preference function from the Preference Ranking Organization Method for Enrichment Evaluation method into a feature selection framework. Our method diverges from traditional single-source focused techniques, providing a tailored solution for the intricacies of multi-source environments, especially those characterized by ordered information. We propose an algorithm that synergizes rough set theory with the preference function to quantitatively assess and rank the relevance of features for decision-making. This is achieved through the construction of preference index matrices, which encapsulate the attribute bias order relationships within the multi-source sequential decision-making information system. To validate the efficacy of our method, we conducted comprehensive experiments on 12 diverse and complex University of California, Irvine public datasets. The results indicate that our feature selection method surpasses existing approaches in terms of accuracy and efficiency, while also providing enhanced insights into the decision-making processes within multi-source ordered information systems.

1. Introduction

Feature selection, which is also called as attribute reduction, plays an major role in the processing and application of big data. As society progresses, the scale and complexity of data have escalated, necessitating the development of more sophisticated feature selection methodologies. More and more feature selection algorithms are being researched in various fields and validated in applications. A Bommert et al. propose filter based feature selection method and applied the method to mathematical calculations in high-dimensional data (Bommert et al., 2020); Rao H, Shi X, Rodrigue A K, et al. propose a novel feature selection method based on bee colony and gradient boosting decision tree (Rao et al., 2019); Chen R C et propose a machine learning-based feature selection method (Chen et al., 2020); MA et al. propose a class-specific feature selection method based on the fuzzy information-theoretic metrics (Ma et al., 2024); Qian's team propose a novel multi-label feature selection algorithm via considering importance of related labels with each sample (Qian et al., 2024) etc. So from above research,it can be obtained that feature selection plays a crucial role in the age and society.

Rough set, building upon the existing relation theory and set theory, have been used to handle the uncertainty computing and incomplete data, which understands knowledge as the partitioning of data. It classify the elements in the set via equivalence relationships. And the equivalence relationship constitutes the division of the data space. The approximate space can be established based on the undistinguished equivalent classes. Under the approximate space, the upper approximation and under approximation are used to approximate a boundary-fuzzy set.

As a pivotal topic within the domain of rough set theory, feature selection serves a dual purpose. It streamlines datasets by eliminating redundant attributes while preserving those with analytical value, thereby enhancing the efficacy of subsequent computational processes. Since the inception of rough set theory by Pawlak (Pawlak and Pawlak, 1998), the field has garnered considerable attention (Modrzejewski, 1993; Swiniarski and Skowron, 2003; Zhong et al., 2001), particularly for its capacity to elucidate feature selection outcomes. As we face increasingly complex data forms and diverse data types, scholars have now proposed more novel feature selection algorithms based on rough sets. About the improvement of rough set, Qiu et al. propose a hierarchical feature selection based on the fuzzy rough set approach using the Hausdorff distance (Qiu and Zhao, 2022); Zhang et al. constructed a feature selection method based on neighborhood rough set for heterogeneous data (Zhang et al., 2022a), Jain et al. even raise a new feature selection method via using intuitionistic fuzzy set theory to get better

* Corresponding author.
E-mail addresses: chxuwh@gmail.com (W. Xu), 3168347551@qq.com (Z. Yang).

https://doi.org/10.1016/j.engappai.2024.109935
Received 19 July 2024; Received in revised form 13 November 2024; Accepted 23 December 2024

effect (Jain et al., 2020). On the different types of data, a matrix-based feature selection method was proposed for ordered data (Xu and Yang, 2023); And one for interval-ordered data (Li et al., 2022) and so on. From previous research, it can be seen that a considerable number of feature selection theory methods have been proposed and studied, and applied to different scenarios for data of different types.

In practical applications, certain information systems exhibit relationships that transcend binary interactions, manifesting instead as preference or advantageous relationships. These systems, termed ordered information systems, encapsulate inherent advantageous relationships among entities. Utilizing these relationships, Scholars have proposed many feature selection methods for ordered data in different data situations: the feature selection for interval-valued ordered data (Li et al., 2022) and for dynamic ordered data (Sang et al., 2020), even for dynamic interval-valued ordered data (Sang et al., 2021). It could be seen that the feature selection of ordered data has been extensively and deeply studied in rough sets, through the hidden partial order relationships in ordered data.

However, in the realm of real-world data, we frequently encounter scenarios that extend beyond single two-dimensional information systems. Often, we are tasked with amalgamating multiple such systems to unearth a trove of valuable information. For instance, comprehensive assessments of institutions like schools or hospitals may integrate diverse perspectives, such as those of a city, region, or patients, underscoring the significance of a holistic approach. Via the author's review on research for multi-source information systems, the main study results focus on information fusion. Xu and Yu came up with a novel information fusion based on triangular fuzzy information granule computing (Xu and Yu, 2017); Che et al. addressed the problem of information fusion and characterization for multi-source information system based on combination of evidence theory, probability theory and information entropy (Che et al., 2018); And a new information infusion model for outliers detection was studied (Zhang et al., 2023); scholars had also reviewed the information fusion related for rough set theory of multi-source information systems (Zhang et al., 2021). A considerable amount of research on multi-source information systems focuses more on information fusion and its application in the context of big data.

But what this paper really wants to stress is not the research on information fusion in multi-source information systems. After reviewing the research on feature selection and multi-source information systems, Current scholarly efforts in multi-source information systems have predominantly focused on information fusion (Xu and Yu, 2017; Che et al., 2018; Zhang et al., 2023, 2021; Guo et al., 2024), with feature selection receiving comparatively less attention, especially in multi-source ordered decision information systems. The existing information fusion methods mostly converted multi-source information systems into single source information systems by extracting important information from different sources (Xu et al., 2022a; Zhang et al., 2022b; Saadi et al., 2018; Guo et al., 2023), where information loss is inevitable. In this paper, all information sources will be considered in feature selection.

For existing feature selection, uncertainty measure is a common method (Yuan et al., 2024, 2023), which is constructed by granularity computing (Jensen, 2008; Anaraki and Eftekhari, 2013) such as feature selection based on weighted neighborhood rough set (Pan et al., 2023), feature selection based on k-nearest-neighbor rough set (Hu et al., 2022) and feature selection using composite entropy based uncertainty (Xu et al., 2022b). Deng et al. found a new feature selection using two-step multi-association information granule and apply it to identify schizophrenia (Ju et al., 2024). And divergence-based fuzzy rough sets was also used in the feature selection (Jiang et al., 2022). Uncertainty is also used in the proposed feature selection algorithm. The uncertainty measure, degree of approximation is used as the evaluation of attributes. Then what is used for feature selection in multi-source ordered decision information system is mainly a decision-making method, Preference Ranking Organization Method for Enrichment Evaluation(PROMETHEE) (Brans et al., 1986).

Since PROMETHEE method was proposed, it have been used in many yields of society (Behzadian et al., 2010): Ranking of sustainable Medical Tourism Destinations was studied by combined fuzzy SWARA-PROMETHEE approach (Ghasemi et al., 2021); Optimization model based on combination of classic mean–variance approach and PROMETHEE method was used for portfolio of palm producers (Ahmadi and Peivandizadeh, 2022); Ghasemi Peiman and Ehsan Talebi Brijani propose a integrated approach which combines the Fuzzy Analytical Hierarchy Process and PROMETHEE for selecting the best flexible Manufacturing System (Ghasemi and Talebi Brijani, 2014); PROMETHEE was used in the GIS-based crisis management too (Choukolaei et al., 2023); Also the PROMETHEE play a role in the incomplete multi-scale data (Deng et al., 2022). It can be seen that PROMETHEE decision-making method have been applied in many aspects of society and many studies focus more on the combination of the PROMETHEE with other theoretical methods (Gul et al., 2018; Bogdanovic et al., 2012). This paper also uses the PROMETHEE to apply it to feature selection, which will be classified in the subsequent paper

From the review for the whole related work and for addressing the issue of few algorithms for multi-source information systems, this paper propose a feature selection algorithm for multi-source ordered decision information system. About the novelty for study of this paper, the paper mainly use the combination of the tool of uncertainty-measure and PROMETHEE to sort attributes for feature selection which are regarded as objects that need decision-making via the construction of attribute evaluation matrix and attributes preference matrix. Specifically, the approximation of rough set is used as an element to construct the attribute evaluation matrix. And PROMETHEE generates the preference matrix based on the attribute evaluation matrix.

The preceding discussion underscores the significance and necessity of feature selection within multi-source ordered information systems. This paper introduces a novel approach to achieve this goal, employing the upper and lower approximations from the rough set model within dominant domains to evaluate and rank features. The ranking is further refined using the PROMETHEE method, which serves as a robust framework for assessing feature quality. The principal contributions of this work are outlined as follows:

(1) The paper use the approximation of attributes to labels as an evaluation of attributes, and use this to assess the relative importance of attributes.

(2) We propose a feature selection method based on the PROMETHEE decision-making method and attribute evaluation matrix for multi-source ordered decision system, which helps addressing the issue of lack of feature selection for multi-source decision information system.

(3) To validate our approach, we conducted experiments on 12 datasets from the UCI Machine Learning Repository. Our method yielded satisfactory outcomes across all datasets, substantiating its effectiveness and generalizability.

The structure of this paper is as follows: Section 2 of the paper mainly describes the preparation of the feature selection method, and makes some definitions of the Ordered Decision Information System (ODIS) and Multi-source Ordered Decision Information System (MSODIS), while Section 3 of the paper mainly describes how the PROMETHEE method can be applied to the feature selection of multi-source ordered information system. The specific process of feature extraction is demonstrated through examples. The Section 4 of the paper gives the performance results of the proposed method on 12 datasets, which verifies the validity of the proposed method in this paper. The Section 5 summarizes the work of this paper and looks forward to the future direction of the work.

2. Preliminaries

In the following sections of this chapter, some definitions and symbols are given for understanding the forthcoming discussion.

Table 1
An ordered decision information system.

U	a_1	a_2	a_3	a_4	d
x_1	0.92	0.90	0.93	0.95	1
x_2	0.82	0.84	0.85	0.86	1
x_3	0.79	0.81	0.82	0.80	1
x_4	0.72	0.73	0.75	0.76	2
x_5	0.68	0.67	0.64	0.69	2

2.1. Multi-source ordered decision information system

A tuple $S = (U, AT \cup d, V, f)$, where $U = \{x_1, x_2, \dots \dots, x_n\}$ is an universe, $AT = \{a_1, a_2, \dots \dots, a_m\}$ is a finite non empty set of attributes and U/d $= \{d_1, d_2, \dots \dots, d_r\}$ is the is a decision partition of U for d, and $U = \cup_{r=1} d_r$. it is generally believed that if the label values of the samples are the same, then these samples belong to the same class.$V = \cup_{a\in AT} V_a$, V_a is the value domain of attribute f: $U \times AT \rightarrow V$ is the function of the information system, where $f(x,a) \in V_a, \forall a \in AT$ and $\forall x \in U$. The tuple S can be defined as a decision information system(DIS).

If a domain of a certain attribute in a DIS is sorted by increasing or decreasing preference, then that attribute is called a criterion. If all attributes are criteria, then the decision information system is called an ordered decision information system(ODIS), it can be expressed as $S_{\geq} = (U, AT \cup d, V, f)$.

In a multi-source ordered decision information system(MODIS), different information sources have different values for U under the evaluation criteria set AT, It can be written as $I_{\geq} = \{U, \{AT : \{a_1^1, a_2^1, \dots, a_j^p, \dots, a_m^s\} p = 1,2,\dots \dots, s, k = 1,2,\dots \dots, m\} \cup d, \{V_k^p : p = 1,2,\dots \dots, s, k = 1,2,\dots \dots, m\}, f_p\}$, Where V_j^p represents the value of the kth attribute under the pth information source, U is an universe, and d is the label attribute of U. $f_p(x, a_j^p) \in V_j^p, \forall a \in AT^p$ and $\forall x \in U$, where AT^p is the attribute set of the pth information source.

Given a $S_{\geq} = (U, AT \cup d, V, f)$, for $\forall F \in AT$ and neighborhood radius parameter β, the dominance neighborhood relation R_F^{β} follow as

$$R_F^{\beta} = \{(x,y) \in U \times U : f(x,a) \geq f(y,a) \wedge f(x,a) - f(y,a) \leq \beta, \forall a \in F\}. \quad (1)$$

Given $S_{\geq} = (U, AT \cup d, V, f)$ as a ODIS, $\forall F \in AT$, $F \neq \emptyset$, the two sets of relationships derived from x are called F-dominating neighborhood sets and F-dominated neighborhood sets, they are written as

$$R_F^{\beta^+}(x) = \{y \in U, y R_F x\}, \quad (2)$$

$$R_F^{\beta^-}(x) = \{y \in U, x R_F y\}. \quad (3)$$

Example 2.1. As shown in Table 1, Table 1 represents an ordered decision information system, where a_1, a_2, a_3, and a_4 represent criteria 1, 2, 3, and 4, and $x_1 - x_5$ represent five samples. And $F = \{a_1, a_2, a_3, a_4\}$, $U = \{x_1, x_2, x_3, x_4, x_5\}$, Here, we take the dominance neighborhood radius parameter $\beta = 0.1$.

Based on the dominance neighborhood radius parameters we have determined, we can get the F-dominance neighborhood sets for each sample: $R_F^+(x_1) = \{x_1\}$, $R_F^+(x_2) = \{x_1, x_2\}$, $R_F^+(x_2) = \{x_2, x_3\}$, $R_F^+(x_4) = \{x_3, x_4\}$, $R_F^+(x_5) = \{x_4, x_5\}$.

$S_{\geq} = (U, AT \cup d, V, f)$ is given as an ODIS, For any $F \in AT, d_n \in d$, The upper and lower approximations of d_n are respectively like this

$$\underline{R_F^{\beta^+}}(d_n) = \{x \in U, R_F^+(x) \in d_n \wedge x \in d_n\}, \quad (4)$$

$$\overline{R_F^{\beta^+}}(d_n) = \{x \in U, R_F^+(x) \wedge d_n \neq \emptyset \wedge x \in d_n\}. \quad (5)$$

So, the degree of approximation for d_n in the attribute set F can be calculated as

$$\gamma_F^{\beta}(d_n) = \frac{|\underline{R_F^{\beta^+}}(d_n)|}{|\overline{R_F^{\beta^+}}(d_n)|}. \quad (6)$$

Example 2.2. From Table 1, $U/d = \{\{x_1, x_2, x_3\}, \{x_4, x_5\}\}$,where $d_1 = \{x_1, x_2, x_3\}, d_2 = \{x_4, x_5\}$,Then the degree of approximation for $d_1 and d_2$ can be obtained :

$$\gamma_F^{\beta}(d_1) = \frac{|\underline{R_F^{\beta^+}}(d_1)|}{|\overline{R_F^{\beta^+}}(d_1)|} = 0.75 \qquad \gamma_F^{\beta}(d_2) = \frac{|\underline{R_F^{\beta^+}}(d_2)|}{|\overline{R_F^{\beta^+}}(d_2)|} = 0.5$$

Multi-source ordered decision information system(MODIS) $I_{\geq} = \{U, \{AT_k^p : p = 1,2,\dots \dots, s, k = 1,2,\dots \dots, m\} \cup d, \{V_k^p : p = 1,2,\dots \dots, s, k = 1,2,\dots \dots, m\}, f\}$ is given. Let $M^p = (m_{ij}^p)_{n*m}$ be the pth information source IF_p, and $V = (v_{ij})_{n*m}$ represents the center of the MODIS, then the consensus degree (Shen et al., 2023) of M^p is defined as

$$WCD_p = 1 - \frac{1}{mn} d(M^p, V). \quad (7)$$

where $d(M^p, V)$ is the Manhattan distance between M^p and V

$$d\left(P^k, V\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} \left| m_{ij}^p - v_{ij} \right|. \quad (8)$$

The calculation for information source center is

$$V = \sum_{i_p \in \mathbb{IN}} \frac{M^p}{|\mathbb{IN}|}. \quad (9)$$

where $|\bullet|$ represents the number of information sources in $\mathbb{IN}$, $\mathbb{IN}$ is the set of all information sources, $p \in R$.

2.2. PROMETHEE method

In this subsection, the following PROMETHEE-methods will be briefly reviewed and how they can be applied to the ODIS.

The PROMETHEE technique, recognized for its prevalence, has garnered extensive application across various decision-making scenarios, yielding commendable empirical outcomes. Central to the PROMETHEE approach is the formulation of a preference framework that delineates the comparative advantage of one option over another, thereby facilitating an informed decision-making process. The brief steps of the method are following

step1: Determine the generalized criteria to be studied.

step2: Establish the preference function P within the context of the generalized criterion.

step3: Weighting the preference functions under different criteria.

step4: Compute the matrix of multi-criteria preference indices.

step5: Determine the outflow and inflow based on the preference indices.

step6: Derive the net flow from the outflow and inflow figures.

step7: A partial order can be deduced using the outflow and inflow data.

step8: A complete order can be established based on the outflow and inflow information.

The PROMETHEE methodology encompasses two distinct approaches, namely PROMETHEE I and PROMETHEE II, corresponding to the processes outlined in Steps 7 and 8, respectively. For the purposes of this paper, the focus is placed on the application of PROMETHEE II.

Since the PROMETHEE II method is ultimately applied to the attribute evaluation matrix defined in Definition 3.1 to compare the attributes in this thesis, there is no need to consider whether Step 1 in Table 2 is real-valued or not, the key to using PROMETHEE II is to define the appropriate preference function P.

It is an ODIS with real-valued information is given. Under the generalized criterion, according to Rao et al. (2019) and Chen et al. (2020), for any two objects, the preference function is as follows

$$P : U \times U \rightarrow (0, 1). \quad (10)$$

Given that the preference function serves to measure the disparity between the assessment values of a pair of entities, the juxtaposition of abject a and object b in the context of a specific criterion a_m can be articulated as follows.

If the criterion a_j is benefit, then

$$H_j = P(f(a,a_j) - f(b,a_j)). \tag{11}$$

If the criterion a_j is benefit, then

$$H_j = P(f(b,a_j) - f(a,a_j)). \tag{12}$$

It is essential to ascertain a decision-maker's preference for one entity, denoted as 'a', in contrast to another, referred to as 'b'. This ascertainment is achieved by aggregating the comparative outcomes of 'a' in relation to 'b' across all evaluative dimensions. Concurrently, it is imperative to quantify the significance of each criterion C_j within the synthesis process, which pertains to the assignment of criterion weights. Should the decision-maker regard all criteria with uniform importance, then the weights can be distributed evenly. Conversely, if the decision-maker perceives the criteria with varying degrees of importance, the weights must reflect this disparity. Therefore, when calculating, we need to consider the weight of each criterion, and the summary formula for the preference index is as follows:

$$\mathbb{P}(a,b) = \sum_{j=1}^{m} H_j \times \omega_j. \tag{13}$$

The aggregate of all benefits that an object holds over every other object is denoted by the term "leaving flow"

$$\eta^+(a) = \sum_{b\in U} \mathbb{P}(a,b). \tag{14}$$

The leaving flow is an aggregate measure of the superiority that an object possesses over every other object across all evaluative criteria.

The sum of all the advantages of other objects relative to one object is given as entering flow

$$\eta^-(a) = \sum_{b\in U} \mathbb{P}(b,a). \tag{15}$$

The entering flow is the cumulative advantage that all other objects have over a particular object across the full spectrum of criteria.

From Eqs. (14) and (15), the net flow is called as the total advantage of an object in the entire decision-making system, which is

$$\eta(a) = \eta^+(a) - \eta^-(a). \tag{16}$$

In this paper, the preference function is defined as a linear preference function, which is

$$P_j(a,b) = \begin{cases} 0 & dis(a,b) \le 0 \\ \frac{dis(a,b)}{dis^j_{max}} & 0 < dis(a,b) < dis^j_{max} \\ 1 & dis(a,b) \ge 1. \end{cases} \tag{17}$$

In the Eq. (17), where

$$dis(a,b) = f(a,a_j) - f(b,a_j), \tag{18}$$

And the dis_{max} is given as following

$$dis^j_{max} = max(f(x,a_j)) - min(f(x,a_j)), \tag{19}$$

And the rule that the preference function is a linear preference function has also been used in subsequent experiments.

3. PROMETHEE-based feature selection for multi-source ordered information systems

In this section, this paper will provide how to evaluate the given attributes based on the relevant definitions in the previous text, and how to rank the attributes to achieve the purpose of feature selection. In general, the whole process of the method is shown in Fig. 1.

3.1. The feature selection for ODIS

In this subsection, some necessary concepts of feature selection method will be defined, and how to specifically perform feature selection on ODIS is presented by giving a example.

Definition 3.1. An ordered decision information system(ODIS) is given as $S_{\geq} = (U, AT \cup d, V, f)$, where $U = \{x_1, x_2, \dots \dots, x_n\}$ is a universe, $AT = \{a_1, a_2, \dots \dots, a_m\}$, $U/d = \{d_1, d_2, \dots \dots, d_r\}$ is the decision partition of U for d, $d = \{D_1, D_2, \dots \dots, D_r\}$ is the set of labels. As shown, the ODIS is a matrix $S_{n\times m}$ with r labels, so the attribute evaluation matrix of the ODIS is matrix $E_{m\times r}$, where $E_{xy} = \gamma^{\beta}_{a_x}(d_y) = \frac{|R^{\beta+}_{a_x}(d_y)|}{|\overline{\overline{R^{\beta+}_{a_x}}}(d_y)|}$, $x = 1, 2, \dots \dots, m; y = 1, 2, \dots, r$.

So according to Definition 3.1, It can obtain an attribute evaluation matrix $E_{m\times r}$ from an ordered information decision-making system $S_{n\times m}$. In the attribute evaluation matrix, the value e_{mr} in the mth row and rth column represents the approximation value of the mth attribute under the rth label and is used as the performance score of the mth attribute under the rth label.

According to Eqs. (13) to (17), afterwards, we can obtain the preference index matrix between attributes, According to Eq. (13), when calculating the preference value of one attribute for another attribute, it is necessary to consider the weight of each class label. We define the label weight as follows, as shown in Eq. (20)

$$\omega_r = \frac{|d_r|}{\sum_{y=1}^{r} |d_y|}. \tag{20}$$

By Eq. (20), it can be concluded that the number of class label samples determines the weight of the label.

Definition 3.2. An ordered decision information system(ODIS) is given as $S_{\geq} = (U, AT \cup d, V, f)$, where $U = \{x_1, x_2, \dots \dots, x_n\}$ is a universe, $AT = \{a_1, a_2, \dots \dots, a_m\}$, $U/d = \{d_1, d_2, \dots \dots, d_r\}$ is the decision partition of U for d, $d = \{D_1, D_2, \dots \dots, D_r\}$ is the set of labels. The attribute evaluation matrix can be obtained as$E_{m\times r}$. From the attribute evaluation matrix, the preference index matrix of ODIS is can be expressed as $\mathbb{P}_{m\times m}$, where $\mathbb{P}_{xy} = \mathbb{P}_{(a_x,a_y)}$according to the Eqs. (13) to (17) for attribute evaluation matrix $E_{m\times r}$.

Then from Eq. (16), the net inflow sequence for the attribute set AT can be calculated as $\eta(AT) =$

$$[\eta^+(a_1) - \eta^-(a_1), \eta^+(a_2) - \eta^-(a_2), \eta^+(a_3) - \eta^-(a_3), \dots \dots, \eta^+(a_m) - \eta^-(a_m)].$$

Next, This paper will provide an example to demonstrate how to calculate the net inflow of each attribute under an ODIS.

Example 3.1. As shown in Table 2, It is a ordered decision information system(ODIS), In this example, $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$, $AT = \{a_1, a_2, a_3, a_4, a_5\}$, $d = \{1,2\}$, $U/d = \{\{x_1, x_2, x_3, x_5\},\{x_3, x_6, x_7\}\}$. Then, According to Eqs. (11) to (17), and make $\beta = 0.25$, the preference index matrix of Table 2 can be obtained in Table 3. Then, by the Eq. (20), the weights of two labels 1 and 2 in Table 2 can also be calculated as $w_{d_1} = \frac{4}{7}, w_{d_1} = \frac{3}{7}$.

After that, the attribute evaluation matrix is obtained in Table 4 according to Definition 3.1 and the Eq. (13). Afterwards, based on the entering inflows and leaving flows proposed in Eqs. (14) through (16) we end up with the preference index matrix with net inflows and net outflows in Table 5. Finally the net inflow for each attribute is obtained according to Eq. (16) as follows

$$\eta = [-0.0616, 2.5918, -0.0616, -0.0616, -2.4079].$$

According to the net inflows in descending order are

$$[a_2, a_1, a_3, a_4, a_5],$$

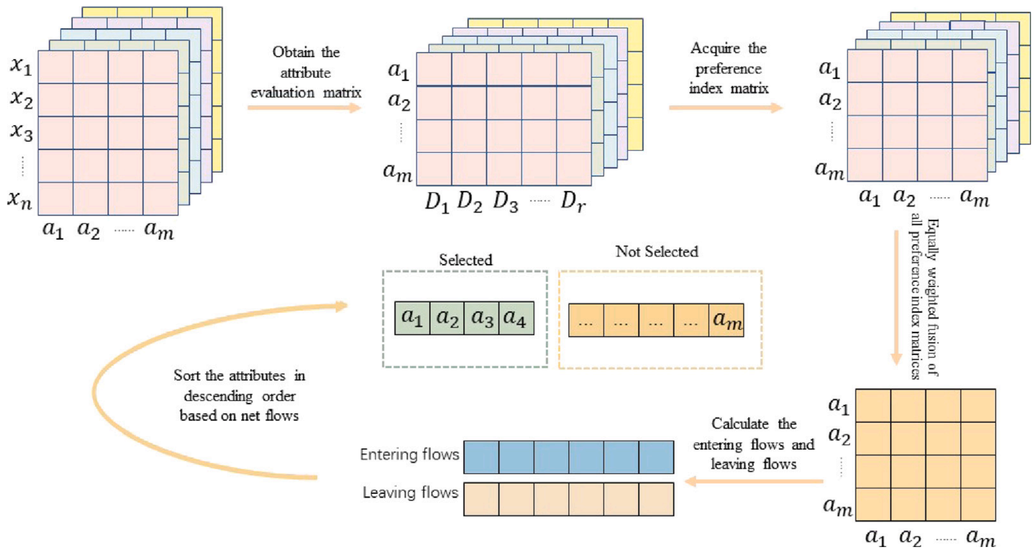

Fig. 1. The process of the feature selection.

Table 2
An ODIS is given.

U	a_1	a_2	a_3	a_4	a_5	d
x_1	0.8571	0.8718	0.2551	0.4227	0.6003	1
x_2	0.9229	0.8954	0.3879	0.9432	0.2178	1
x_3	0.7013	0.5007	0.3607	0.2257	0.5482	1
x_4	0.8242	0.9239	0.7609	0.1821	0.6731	2
x_5	0.4926	0.5887	0.9059	0.9469	0.4217	1
x_6	0.1075	0.3597	0.5401	0.3826	0.1103	2
x_7	0.7149	0.1034	0.7537	0.5618	0.5388	2

Table 3
Attribute evaluation matrix of Table 2.

U	d_1	d_2
a_1	0.33	0.20
a_2	0.40	0.40
a_3	0.33	0.20
a_4	0.33	0.20
a_5	0.16	0.16

where a_2 is the attribute with the best approximation, followed by $a_1, a3, a4$ and the worst approximation is a_5After that, according to the need of selecting the number of attributes, starting from attribute $a5$, the number of attributes that satisfy the quantity requirement is selected, and then the purpose of feature extraction is achieved.

With Example 3.1, the process of feature extraction for a single source has been described clearly, essentially this paper uses the degree of approximation to give a score to the attributes and then uses a preference function to compare the attributes with each other and finally uses the net inflow of each attribute as the basis for ranking the attributes.

3.2. The feature selection for MODIS

According to Example 3.1, the process of ODIS feature selection for a single source has been described clearly, this paper mainly focuses on how to carry out MODIS feature selection under the condition of multi-source, this subsection will be based on the feature selection method for a single ODIS in the previous section to describe how to carry out MSODIS feature selection under the condition of multi-source.

Table 4
The preference index matrix of Table 2.

U	a_1	a_2	a_3	a_4	a_5
a_1	0	0	0	0	0.4693
a_2	0.5306	0	0.5306	0.5306	1
a_3	0	0	0	0	0.4693
a_4	0	0	0	0	0.4693
a_5	0	0	0	0	0

Table 5
The preference index matrix of Table 2.

U	a_1	a_2	a_3	a_4	a_5	η^+
a_1	0	0	0	0	0.4693	0.4693
a_2	0.5306	0	0.5306	0.5306	1	2.5918
a_3	0	0	0	0	0.4693	0.4693
a_4	0	0	0	0	0.4693	0.4693
a_5	0	0	0	0	0	0
η^-	0.5306	0	0.5306	0.5306	2.4079	

Definition 3.3. A multi-source ordered decision information system(MODIS) is given as $I_{\geq} = \{U, \{AT_k^p : p = 1, 2, \ldots \ldots, s, k = 1, 2, \ldots \ldots, m\} \cup d, \{V_k^p : p = 1, 2, \ldots \ldots, s, k = 1, 2, \ldots \ldots, m\}, f_p\}$, Where V_k^p represents the value of the kth attribute under the pth information source, U is a universe, and d is the label attribute of $U.AT_k^p = \{a_1^1, a_2^1, a_3^1, \ldots., a_k^p, \ldots, a_{m-1}^s, a_m^s\}$, a_k^p represents the kth attribute under the pth source, It is worth noting that the division of class labels U/d by object U is the same under any source. For the MOIDS, $\mathbb{E}_{m\times r}$ are the attribute evaluation matrices. For each ordered decision-making information system an attribute evaluation matrix is obtained according to Definition 3.1, so the attribute evaluation matrices for MODIS can be written as $\mathbb{E}_{m\times r} = \{E_{m\times r}^1, E_{m\times r}^2, \ldots \ldots, E_{m\times r}^s\}$, where $E_{m\times r}^s$ is the attribute evaluation matrix under the pth source.

Definition 3.4. For multi-source ordered decision information system(MODIS), we define its pth source as IF_p, and IF is a single ODIS,in other words MODIS is set of multiple ODIS that can be written as $\{IF_1, IF_2, \ldots \ldots, IF_p, \ldots \ldots, IF_s\}$, where IF_p represents the pth information source.

In other words, a preference index matrix between attributes is obtained in each source, and how to take into account the information of all sources to obtain a final preference index matrix is an important issue.

Table 6
An example of MODIS is given.

U	a_1			a_2			a_3			a_4			d
	a_1^1	a_1^2	a_1^3	a_2^1	a_2^2	a_2^3	a_3^1	a_3^2	a_3^3	a_4^1	a_4^2	a_4^3	
x_1	0.30	0.11	0.67	0.26	0.61	0.70	0.05	0.13	0.61	0.54	0.24	0.30	2
x_2	0.47	0.32	0.20	0.63	0.85	0.19	0.97	0.66	0.79	0.90	0.54	0.29	1
x_3	0.91	0.29	0.65	0.52	0.73	0.30	0.52	0.39	0.14	0.18	0.80	0.40	2
x_4	0.95	0.25	0.31	0.42	0.56	0.24	0.86	0.86	0.58	0.67	0.22	0.24	1
x_5	0.62	0.40	0.74	0.27	0.31	0.72	0.89	0.38	0.69	0.20	0.95	0.10	2
x_6	0.40	0.84	0.94	0.99	0.97	0.50	0.73	0.38	0.89	0.44	0.95	0.19	1
x_7	0.56	0.44	0.59	0.41	0.67	0.96	0.72	0.82	0.99	0.40	0.89	0.21	2

Table 7
The attribute evaluation matrices for Table 6.

U	d_1			d_2		
	i_1	i_2	i_3	i_1	i_2	i_3
a_1	0.20	0.20	0.75	0.33	0.33	0.75
a_2	0.60	0.28	0	0.50	0	0.42
a_3	0.67	0.20	0	0.80	0.33	0.28
a_4	0.33	0.33	0	0.20	0.20	0.28

Definition 3.5. $I_{\geq} = \{U, \{AT_k^p : p = 1, 2, \dots\dots, s, k = 1, 2, \dots\dots, m\} \cup d, \{V_k^p : p = 1, 2, \dots\dots, s, k = 1, 2, \dots\dots, m\}, f_p\}$ is given as a MODIS. The preference index matrices can be written as $\mathbb{Pre} = \{Pre_1, Pre_2, \dots\dots, Pre_s\}$. Here Pre_p is the preference index matrix of the pth information.

In order to get a comprehensive preference index matrix between attributes of multi-source information system, all sources will have a weight, and the final preference index matrix between attributes satisfies the following equation

$$\mathbf{Pre} = \sum_{p=1}^{s} \omega_p \times Pre_p, \tag{21}$$

where ω_p is the weight of pth information source, Pre_p represents the preference index matrix of pth information source.

In this paper we consider that the information contained in all sources is of equal importance, in other words, all sources have equal weights, i.e., $\omega_p = \frac{1}{s}$ in Eq. (21), s is the number of information sources.

Afterwards, using the same method applied to the single information system ODIS, i.e., the net flow of each attribute is calculated according to Eqs. (14) to (16) to obtain the sequence of runoffs for all attributes under the multi-source ordered decision information system(MODIS), which is as following $\eta(AT) =$

$$[\eta^+(a_1) - \eta^-(a_1), \eta^+(a_2) - \eta^-(a_2), \eta^+(a_3) - \eta^-(a_3), \dots\dots\dots, \eta^+(a_m) - \eta^-(a_m)].$$

The same as the ODIS, sort the attributes according to the net flow from the largest to the smallest order to get the attribute sequence, and finally select a certain number of attributes according to the feature selection rate, and then complete the feature selection under the multi-source ordered decision information system(MODIS).

The above is the whole process of feature selection under multi-source decision information system. Next, this paper will give a concrete example to illustrate how this method is used for feature selection from a multi-source information system.

Example 3.2. Table 6 is given as a MODIS, As shown in Table 6, which has 4 attributes and 3 information sources a_j^p represents the jth attribute under the pth information source. According to the Definition 3.3, the multi-source attribute evaluation matrices for Table 6 can be obtained, which is shown in Table 7. Next from Table 7 for the four attributes of the three information sources we can get the preference index matrix for each of the three information sources, which is shown in Table 8, Table 8 shows the preference index matrix for each attribute for different information sources. From Eq. (21), then, the final preference index matrix between attributes in Table 6 can be obtained from Table 8, which combines the information contained in each information source and shown in Table 9. Finally, It can be calculated from attribute a_1 to attribute a_4 the net flow of each attribute according to Eq. (16).

$[0.50851212, -0.28351067, 0.33919995, -0.56420141]$.

After that, the attribute sequences are obtained based on the attribute net flows sorted from largest to smallest as follows

$[a_1, a_3, a_2, a_4]$.

Finally the attributes are selected according to the number of attributes to be selected to complete the feature selection. If the feature selection rate is 0.5, which means the features selected in the first 50 percent of the number of attributes, i.e. a1 and a3 are the attributes selected.

The above is the whole process of feature selection for MODIS, which focuses on the calculation of the upper and lower approximation, the approximation is the evaluation of the attributes of this thesis is good or bad indicators.

It is worth noting that, in the specific process, the number of objects in each class will have a greater impact on the calculation of the upper and lower approximations, for example, if the number of objects in a class label is 1, then its approximation can only be calculated as 1, it is clear that this way of evaluating the attributes is biased.

Next, we summarize the process of the feature selection method for MODIS in the Algorithm 1

Algorithm 1 is a static algorithm based on the PROMETHEE method for feature selection of MODIS. The time complexity of the main steps of the algorithm is analyzed next: the time complexity for calculating the attribute evaluation matrix is $O(s \times r \times m)$, the time complexity for calculating the preference index matrix is $O(s \times m^2)$, the time complexity for calculating net flow is $O(s)$, the time complexity for selecting features is $O(\alpha \times s)$. So the total time complexity of Algorithm 1 is $O(s \times m^2)$.

4. Experimental analysis

In this section a series of experiments are held to test the performance of our feature selection method and to compare the FSPA-MODI with other algorithms. The datasets used to conduct the experiments from the UCI (https//archive.ics.uci.edu/datasets). Website are shown in Table 10. The feature selection method in this paper was written by python 3.7 in the environment of Pycharm 2022 and was run on a 3.30 GHz AMD Ryzen 5 5600H with Radeon Graphics, 16.0 GB of memory, 64-bit Windows 11 computer.

Data processing: To confirm the datasets that adopted could represent the multi-source ordered information decision system, some operation about data processing will be described in the following: (1): Firstly, all datasets will exclude non numeric features and retain attributes that can be sorted in the domain.

(2): Make maximum minimum normalization on all datasets

(3): Since the dataset for this experiment is a single two-dimensional table, this paper constructs the sources by adding white noise for 50% data to combine the multi-source information systems, the specific con-

Table 8
Multi-source preference index matrix for Table 6.

U	a_1			a_2			a_3			a_4		
	i_1	i_2	i_3	i_1	i_2	i_3	i_1	i_2	i_3	i_1	i_2	i_3
a_1	0	0	0	0	0.5714	0.8241	0	0	1	0.1269	0.2285	1
a_2	0.5260	0.2755	0	0	0	0	0	0.2755	0.1758	0.5306	0	0.1758
a_3	0.8730	0	0	0.3469	0.5714	0	0	0	0	0.8775	0.2285	0
a_m	0.1224	0.4285	0	0	0.4959	0	0	0.4285	0	0	0	0

Table 9
The final preference index matrix of Table 6.

$\mathbb{P}$	a_1	a_2	a_3	a_m	$\eta^+(a)$
a_1	0	0.4652	0.3333	0.4518	1.2503
a_2	0.2671	0	0.1504	0.2354	0.6531
x_3	0.2910	0.3061	0	0.3687	0.9658
x_4	0.1836	0.1653	0.1428	0	0.4918
$\eta^-(a)$	0.7418	0.9366	0.6266	1.0560	

struction method can be referred to Eq. (22). Specifically, white noise is added to 50% of the dataset each time to construct a information source.

$$F_i(x,a) = \begin{cases} f(x,a) + n_i, & \text{if } x \text{ in } U_{radom} \\ f(x,a), & \text{else} \end{cases}, \tag{22}$$

where, U_{radom} is the set of 50% dataset, n_i obey a normal distribution with a mean of 0 and a variance of 0.1.

Parameters set: This paper sets two parameters: attribute selection rate α, that is, how many attributes are selected after the attributes are sorted, and we decide $|AT| \times \alpha$ attributes are selected according to attribute selection rate α; Neighborhood radius β, which is used to calculate the approximation degree in Eq. (6). Set β from 0.05 to 0.5 by step 0.5 and for each β set α from 0.1 to 1 by step 0.1. However, for datasets with number of attributes less than 10, set α from $\frac{1}{|AT|}$ to 1 by step $\frac{1}{|AT|}$

Obtaining classification accuracy: this paper classifies the feature-selected dataset by three classifiers: KNN, SVM and DT, and obtains its mean accuracy under the 5-flod cross validation.

Selection of information source: There is another problem to pay attention to that is how to classify after the feature selection, because this paper deals with multi-source information system rather than a single two-dimensional information system, which would not be classified by classifiers. So in the experiments will take into account the problem, according to Eqs. (7)–(10) this paper in the experiments through the introduction of the consensus degree to select the best information source from multiple sources to follow up the classification accuracy test part when the feature selection is complete.i.e

$$IF_{best} = argmax_{IF_p} WCD_p, \tag{23}$$

where IF_{best} is the best information source in the multi information system, and IF_p is the pth information source.

From Eq. (23), the feature-selected IF_p will be classified by the three classifiers, and acquire the accuracy which will be evaluation for feature selection.

4.1. Classification effect of KNN classifier for features-selected dataset

In this subsection, based on the determination of the parameters in the previous section, this paper conducts KNN classification for the MODIS after feature selection, and obtains the classification accuracy at different β and different attribute selection rates α, which is specifically shown in each subfigure in Fig. 2. Through the observation of each subgraph in Fig. 2, it can be found that most of the datasets have high sensitivity to the change of parameters, and the change process of β from 0.05 to 0.5 in some datasets does not have much change in its classification accuracy, which indicates that β should be set according

Algorithm 1: An feature selection PROMETHEE II-based algorithm for multi-source ordered decision information(FSPA-MODI)

Input: MSODIS:$I_{\geq} = \{U, \{AT : \{a_1^1, a_2^1, ..., a_j^p, ..., a_m^s\} p = 1,2,......,s, k = 1,2,......,m\} \cup d, \{V_k^p : p = 1,2,......,s, k = 1,2,......,m\}, f_p\}$,feature selection rate$\alpha$,neighborhood radius$\beta$,U/d $= \{d_1, d_2, ..., d_r\}$
Output: features that selected$AT_{selected}$

$EVA \leftarrow \{\}$
for $p = 1$ *to* s **do**
  create zero matrix :$E_{m\times r}$;
  for $t = 1$ *to* r **do**
    for $j = 1$ *to* m **do**
      Calculate the degree of approximation $e_{jt} \leftarrow \gamma_{a_j}^{\beta}(d_t) = \frac{|\underline{\underline{R_F^{\beta+}}}(d_t)|}{|\overline{\overline{R_F^{\beta+}}}(d_t)|}$;
    end
  end
  $EVA \leftarrow EVA \cup \{E_{m\times r}\}$;
end
$Pre \leftarrow \{\}$;
for $i = 1$ *to* s **do**
  create zero matrix :$P_{m\times m}$;
  for $j = 1$ *to* m **do**
    for $t = 1$ *to* m **do**
      Calculate the preference index of the j-th attribute; for the t-th attribute and use it as the j-th row andt-th column element in the matrix: $p_{jt} \leftarrow \mathbb{P}(a_j, a_t)$;
    end
  end
  $Pre \leftarrow Pre \cup \{P_{m\times m}\}$;
end
calculate the final preference index matrix :$\mathbb{P}_{m\times m} = \sum_{p=1}^{s} Pre[p] \times \omega_p$,where $\omega_p = \frac{1}{p}$;
$\eta_{list} \leftarrow \{\}$;
for *each* $j = 1$ *to* s **do**
  Calculate the net flow of attributea_j; $\eta_{list} \leftarrow \eta_{list} \cup \{\eta(a_j)\}$;
end
Arrange the η_{list} in descending order to get a new sequence of attributes:$AT = \{a_1, a_2, a_3, ..., a_m\}$;
for $p = 1$ *to* $\alpha \times s$ **do**
  $AT_{selection} \leftarrow AT_{selection} \cup AT[p]$;
end
return$AT_{selection}$;

to the size of the change of data in the dataset to set up the range of change of β specifically for a better effect. Overall, it is necessary to find a suitable combination of parameters in the process of feature selection.

The classification accuracy under different β at half of the feature selection rate and the classification accuracy without feature selection are shown in Table 11, from which it can be seen that even with half of the features selected, most of their classification accuracies under

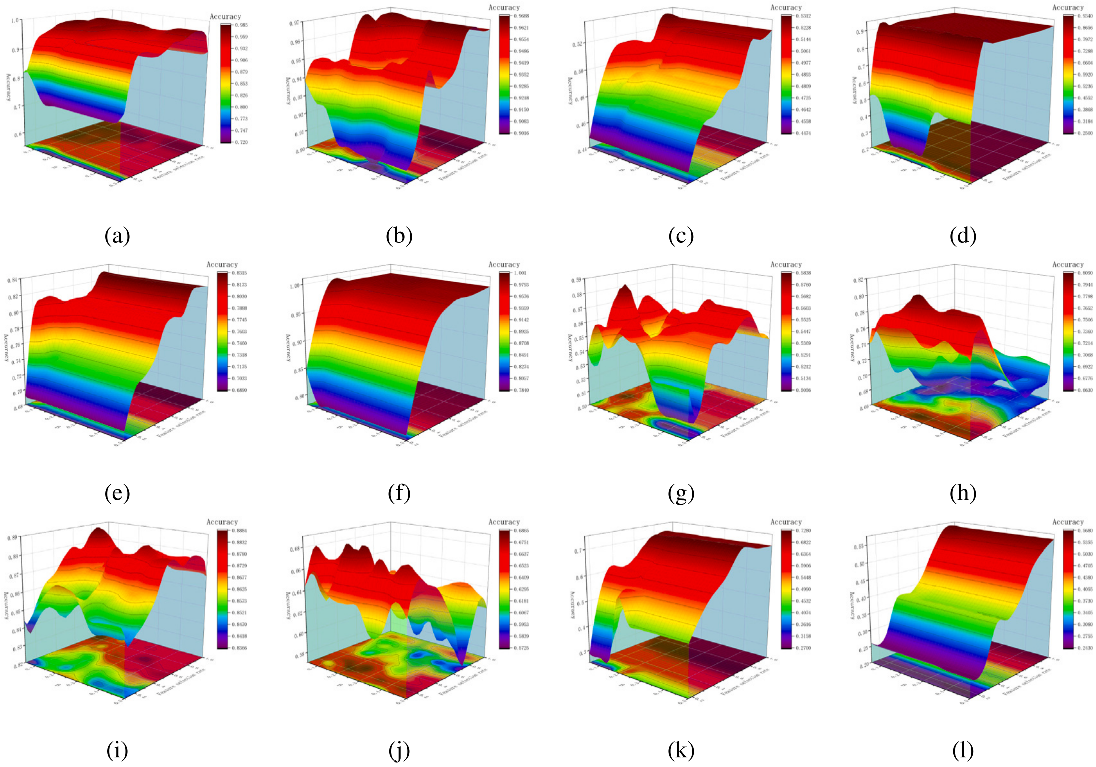

Fig. 2. The accuracy of different β and α: (a) Wine, (b) Breast Cancer, (c) Abalone, (d) Dry Bean, (e) Magic gamma relescope, (f) Occupancy Detection, (g) Hill valley, (h) DARWIN, (i) Parkinson's Disease Classification, (j) Toxicity, (k) Maternal health risk, (l) Yeast.

Table 10
The summary of datasets

No.	Datasets	Samples	Attributes	Classes
Data1	Toxicity	171	1204	2
Data2	DARWIN	174	452	2
Data3	Wine	178	13	3
Data4	Breast Cancer	569	30	2
Data5	Hill valley	606	100	2
Data6	Parkinson's Disease Classification	756	755	2
Data7	Maternal health risk	1014	6	3
Data8	Yeast	1484	8	10
Data9	Abalone	4177	8	3
Data10	Dry Bean	13 611	16	7
Data11	Magic gamma relescope	19 020	10	2
Data12	Occupancy detection	20 560	5	2

the appropriate β are higher than the classification accuracy of the raw data. In some datasets, the highest classification accuracy is lower than that of the raw data under different values of β, such as Data11 and Data12, but the difference between the two is very small.

From the data in Table 11, it can be seen that, on the one hand, Algorithm 1 does select effective features in each dataset, and most of the classification accuracies under half of the features are higher than those of the raw data. On the other hand, the classification accuracies under different β are not the same, which again shows that Algorithm 1 is highly sensitive to the parameters, and it is necessary to choose the appropriate parameters β and α when making feature selection.

4.2. Classification effect of SVM classifier for features-selected dataset

The classification accuracy on SVM after feature selection is shown in Fig. 3, which is the same as the results of KNN classification, most of the datasets have high sensitivity to parameter changes, and some datasets have no significant change in classification accuracy when β changes from 0.05 to 0.5. Overall, it is necessary to find a suitable combination of parameters in the process of feature selection.

The classification accuracy of half of the features selected for each β under SVM is also listed as shown in Table 12, which is different from the KNN classification results, and the results of SVM classification show that there are more datasets with better classification results than those after feature selection, but the classification accuracy of the data after feature selection is very small compared to that of the raw data in this case. Also, there are some case that the classification accuracy of feature-selected data is better than raw data.

The results from Table 12 show that under the SVM classification results, the features that are effective under multiple sources of information in each dataset are indeed selected by Algorithm 1.

4.3. Classification effect of DT classifier for features-selected dataset

The classification accuracy on DT after feature selection is shown in Fig. 4, which is the same as the results of KNN and SVM classification, most of the datasets have high sensitivity to parameter changes, and some datasets have no significant change in classification accuracy when β changes from 0.05 to 0.5. Overall, it is necessary to find a suitable combination of parameters in the process of feature selection.

In Table 13, the classification accuracy of DT under half of the attribute selection rate under different β and the classification accuracy of the raw data are listed, in fact, through the results of Table 13 show that DT's classification results of the data after feature selection are largely the same as SVM, there will be some datasets that will be better classified on the raw data, but in this case, the classification accuracy of the attributes that are selected only half of the attributes is actually very similar to that of the with the raw data is very small.

The classification results of DT are consistent with those of KNN and SVM, which show that Algorithm 1 can effectively select the attributes with better classification effect, while on the other hand, Algorithm 1 has parameter sensitivity, and it is necessary to determine the appropriate parameters.

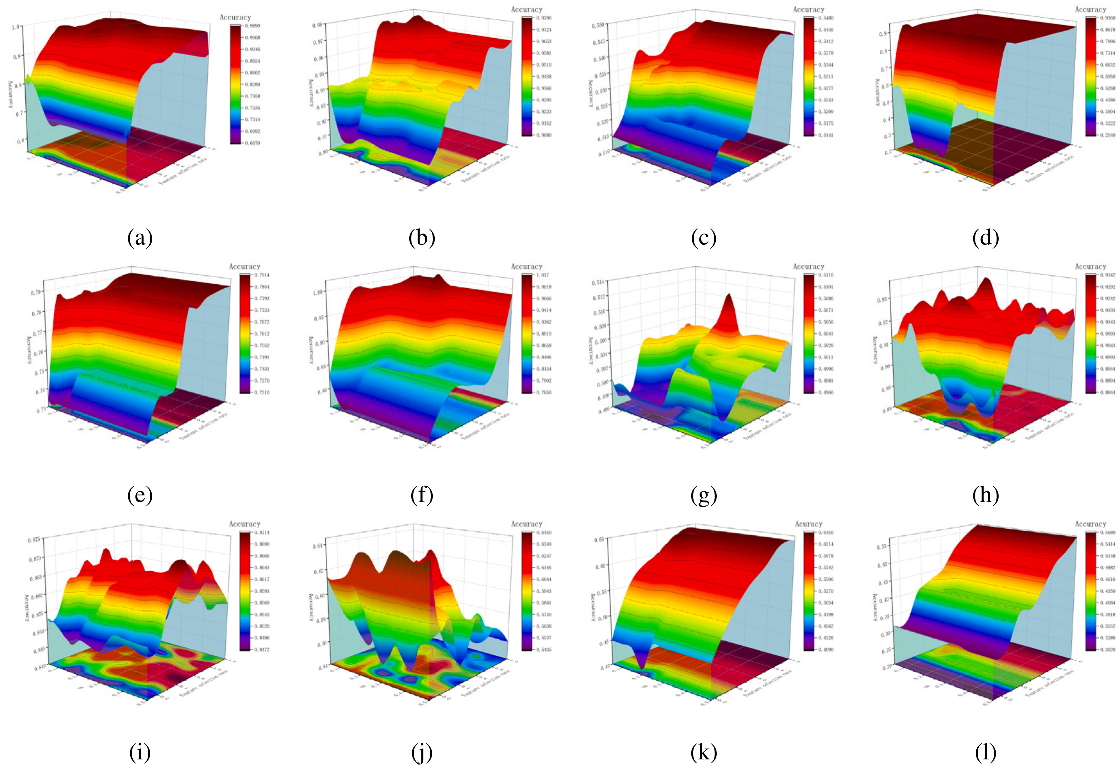

Fig. 3. The accuracy of different β and α: (a) Wine, (b) Breast Cancer, (c) Abalone, (d) Dry Bean, (e) Magic gamma relescope, (f) Occupancy Detection, (g) Hill valley, (h) DARWIN, (i) Parkinson's Disease Classification, (j) Toxicity, (k) Maternal health risk, (l) Yeast.

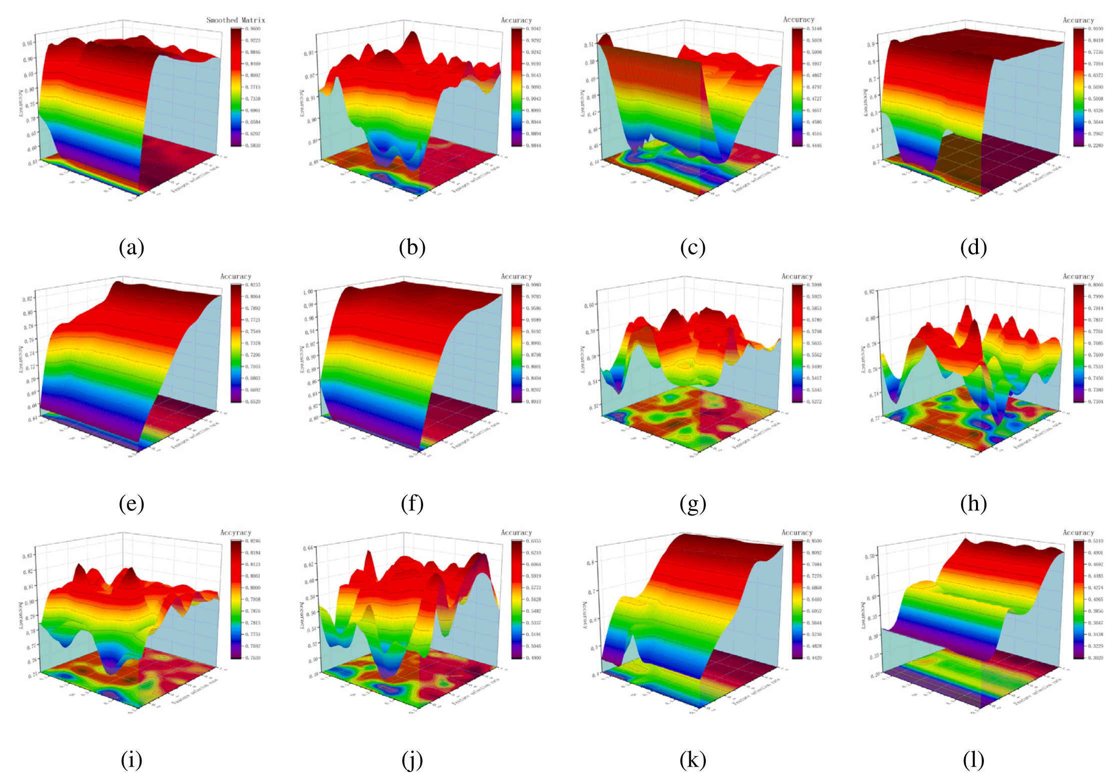

Fig. 4. The accuracy of different β and α: (a) Wine, (b) Breast Cancer, (c) Abalone, (d) Dry Bean, (e) Magic gamma relescope, (f) Occupancy Detection, (g) Hill valley, (h) DARWIN, (i) Parkinson's Disease Classification, (j) Toxicity, (k) Maternal health risk, (l) Yeast.

Table 11
The classification accuracy (%) of feature-selected datasets on KNN.

Dataset	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	05	Raw
Data1	**66.06**	62.55	61.37	61.96	63.14	61.36	60.80	61.98	60.80	61.37	61.94
Data2	69.56	67.84	66.70	68.99	70.13	**72.99**	72.42	71.84	70.70	68.9	68.42
Data3	97.17	**98.31**	95.50	97.73	93.25	93.25	94.36	94.36	94.36	94.36	96.63
Data4	95.94	95.94	94.88	94.18	94.35	94.88	94.88	94.88	94.88	94.88	**96.47**
Data5	55.60	**57.42**	53.63	56.43	56.27	56.60	56.27	56.27	56.27	56.27	55.11
Data6	88.08	86.50	86.90	86.63	**88.48**	87.69	87.69	88.09	87.82	88.22	87.43
Data7	63.30	71.49	71.49	**72.38**	72.38	72.38	72.38	72.38	72.38	72.38	70.60
Data8	41.67	41.67	41.67	41.67	41.67	41.67	41.67	41.74	41.74	41.74	**56.03**
Data9	**51.34**	50.91	49.76	50.95	49.47	49.47	49.47	49.47	49.47	49.47	52.72
Data10	**91.72**	90.55	90.57	90.57	90.57	90.57	90.16	90.16	90.16	90.16	91.35
Data11	80.24	80.24	79.60	79.60	80.2	80.24	80.24	80.24	80.24	80.24	**82.77**
Data12	99.21	98.14	98.14	98.14	98.14	98.14	98.14	98.14	98.14	98.14	**99.23**

Table 12
The classification accuracy (%) of feature-selected datasets on SVM.

Dataset	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	05	Raw
Data1	56.73	58.48	60.25	58.48	**61.42**	60.21	61.41	60.23	58.50	59.07	55.56
Data2	93.29	91.36	91.89	90.48	91.89	91.36	92.24	91.71	92.41	92.94	92.07
Data3	97.19	96.63	95.52	97.76	93.26	93.26	93.25	93.25	93.25	93.25	97.74
Data4	97.18	**97.35**	95.59	94.36	95.06	95.06	94.53	94.53	94.53	94.53	96.82
Data5	50.16	**50.49**	50.0	50.16	50.33	50.33	50.33	50.33	50.33	50.33	50.49
Data6	85.97	86.76	86.37	85.97	87.03	85.84	86.50	86.77	86.37	86.63	85.71
Data7	60.05	63.80	63.80	60.54	60.54	60.54	60.54	60.54	60.54	60.54	**64.09**
Data8	45.24	45.24	45.24	45.24	45.24	45.24	45.24	45.24	45.24	45.24	**56.70**
Data9	53.73	53.92	53.85	54.23	54.14	54.14	54.14	54.14	54.14	54.14	**54.57**
Data10	92.03	91.43	91.36	91.36	91.36	91.36	91.14	91.14	91.14	91.14	**92.10**
Data11	78.46	78.46	78.86	78.86	78.46	78.46	78.46	78.46	78.46	78.46	**79.13**
Data12	98.77	83.24	83.24	83.24	83.24	83.24	83.24	83.24	83.24	83.24	**98.83**

Table 13
The classification accuracy (%) of feature-selected datasets on DT.

Dataset	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	05	Raw
Data1	**62.55**	54.90	59.59	61.36	59.02	60.78	59.07	56.67	61.36	61.94	60.20
Data2	**82.01**	80.55	79.76	80.29	80.02	81.08	79.23	80.42	79.63	78.83	80.82
Data3	90.98	90.44	**93.84**	93.84	92.71	92.69	92.15	92.71	92.71	91.03	90.46
Data4	**93.29**	91.36	91.89	90.48	91.89	91.36	92.24	91.71	92.41	92.94	92.07
Data5	57.09	57.42	55.44	59.40	56.43	**60.06**	58.24	59.24	57.92	58.08	58.90
Data6	**82.01**	80.55	79.76	80.29	80.02	81.08	79.23	80.42	79.63	78.83	80.82
Data7	69.62	78.99	79.38	80.07	80.46	80.66	80.37	80.56	80.56	80.46	**84.51**
Data8	38.57	38.90	39.04	39.11	39.51	39.11	38.90	39.17	38.91	38.70	**50.50**
Data9	47.55	48.13	45.92	47.94	48.13	47.43	47.67	47.50	47.53	47.74	**49.90**
Data10	89.39	88.31	88.90	88.92	88.82	88.97	88.17	88.29	88.16	88.31	**89.50**
Data11	78.51	78.56	77.97	78.05	78.59	78.34	78.60	78.54	78.29	78.53	**81.79**
Data12	98.99	97.33	97.29	97.39	97.26	97.32	97.33	97.33	97.41	97.32	**99.13**

4.4. Comparison of experiments

In this subsection, because of few studies of feature selection for multi-source information, the comparison of Algorithm 1 (FSPA-MODI) with 3 current feature selection methods for single-source ordered information: HAR (Xu and Yang, 2023), HFS-IVO (Sang et al., 2021), WD-HAR (Pan et al., 2023) and 1 fusion model FS-MS (Xu et al., 2022a) was conducted. But some work is done before comparison to make sure the 4 algorithms could be applied on the multi-source ordered information system. Here the more details will be classified about the designing of comparison:

(1): The parameters λ and radius of neighborhood δ of WD-HAR algorithm are set to 0.25 and −0.1 partly.

(2): The parameter p of HFS-IVO algorithm are set as 0.5.

(3): The parameter α and β of FS-MS algorithm are set as 0.4 and 0.5 partly.

(4): The parameter β of FSPA-MODI is set as 0.25 while the attribution reduction rate α is set two values: 30%, 50%, 70% respectively,namely FSPA-MODI(30%), FSPA-MODI(50%) and FSPA-MODI (70%)

(5): To illustrate the effectiveness of FSPA-MODI and other 4 algorithms, the datasets listed in Table 10 are used by the algorithms after the process at the beginning of this section for the datasets. The classification accuracy of the datasets after feature selection under three classifiers: KNN, SVM, DT, as well as the running time of our algorithm with four other algorithms. The mean value of 5-fold cross validation for all classifiers is used as final accuracy.

(6): Attention. For information fusion algorithm FS-MS, the classification accuracy of the fused data will be added into comparison with other algorithms

Firstly, the accuracy of three classifiers on the feature-selected datasets under all algorithms is analyzed. The more detailed data are shown in Tables 14–16. From Tables 14–16, it can be obtained that the performance of FSPA-MOID on the most features-selected datasets is better than other algorithms. For the whole 12 datasets on 3 classifiers, there are 36 cases and HAR, WD-HAR, HFS-IVO, FS-MS, FSPA-MOID perform best on the 1, 2, 5, 1, 21 cases. Also FSPA-MOID performance best on the average accuracy. Generally speaking, from all accuracy of features-selected datasets on three classifiers, FSPA-MOID is still best.

Then the run time of all algorithms is also listed in Table 17. Some explanations about running time must be given. Because of the comparison with algorithms for single source ordered information. The running time of FSPA-MOID only calculated for one information source in the MSODIS. In Table 17, the run time of FSPA-MOID is minimal on

Table 14
The classification accuracy (0 – 1) of feature-selected datasets by reduction algorithms on KNN.

Dataset	HAR	WD-HAR	HFS-IVO	FS-MS	FSPA-MODI (30%)	FSPA-MODI (50%)	FSPA-MODI (70%)	Raw data
Data1	0.613	0.637	0.626	0.62	0.627	0.625	**0.648**	0.643
Data2	0.654	0.625	0.591	0.55	0.747	**0.735**	0.689	0.614
Data3	0.887	0.945	0.787	0.714	0.859	0.87	**0.96**	0.938
Data4	0.948	**0.968**	0.959	0.861	0.915	0.927	0.917	0.964
Data5	0.523	0.527	0.528	0.486	0.475	0.478	0.481	**0.551**
Data6	0.742	0.768	0.800	0.677	0.828	**0.835**	0.821	0.784
Data7	0.633	**0.725**	0.633	0.489	0.419	0.571	0.658	0.723
Data8	0.347	0.497	0.383	0.285	0.3774	0.469	**0.54**	0.521
Data9	0.489	0.517	0.509	0.427	0.466	0.48	**0.519**	0.519
Data10	0.668	0.723	0.761	0.138	0.865	0.902	**0.902**	0.827
Data11	0.825	**0.833**	0.763	0.693	0.735	0.782	0.791	0.824
Data12	0.748	0.859	0.925	0.81	0.755	0.893	**0.944**	0.922
Average	0.673	0.721	0.722	0.721	0.672	0.713	**0.739**	0.735

Table 15
The classification accuracy (0 – 1) of feature-selected datasets by reduction algorithms on SVM.

Dataset	HAR	WD-HAR	HFS-IVO	FS-MS	FSPA-MODI (30%)	FSPA-MODI (50%)	FSPA-MODI (70%)	Raw data
Data1	0.672	0.6724	0.479	**0.672**	0.607	0.555	0.567	0.532
Data2	0.717	0.706	0.757	0.551	0.827	0.816	**0.833**	0.786
Data3	0.904	0.938	0.787	0.769	0.871	0.888	**0.96**	0.956
Data4	0.964	**0.973**	0.964	0.883	0.922	0.926	0.933	0.968
Data5	0.498	0.504	0.499	0.496	0.498	0.509	**0.511**	0.504
Data6	0.784	0.843	0.824	0.746	0.855	0.861	**0.862**	0.808
Data7	0.552	0.592	0.552	0.453	0.462	0.559	0.612	**0.64**
Data8	0.398	0.458	0.452	0.349	0.317	0.397	0.511	**0.559**
Data9	0.534	0.539	0.535	0.511	0.509	0.511	**0.541**	0.495
Data10	0.782	0.834	0.858	0.535	0.864	0.912	**0.913**	0.861
Data11	0.79	0.789	0.755	0.717	0.748	0.781	**0.798**	0.791
Data12	0.794	0.974	**0.987**	0.987	0.768	0.837	0.838	0.986
Average	0.697	0.726	0.739	0.732	0.687	0.712	**0.740**	0.740

Table 16
The classification accuracy (0 – 1) of feature-selected datasets by reduction algorithms on DT.

Dataset	HAR	WD-HAR	HFS-IVO	FS-MS	FSPA-MODI (30%)	FSPA-MODI (50%)	FSPA-MODI (70%)	Raw data
Data1	0.549	0.553	**0.62**	0.561	0.567	0.543	0.52	0.596
Data2	0.708	0.626	0.722	0.534	0.741	**0.77**	0.712	0.699
Data3	0.904	0.876	0.826	0.658	0.876	0.809	**0.904**	0.887
Data4	**0.943**	0.882	0.936	0.822	0.867	0.903	0.894	0.919
Data5	0.547	**0.572**	0.518	0.534	0.524	0.531	0.553	0.553
Data6	0.73	0.752	0.769	0.693	**0.794**	0.789	0.761	0.75
Data7	0.686	0.804	0.686	0.541	0.455	0.592	0.681	**0.805**
Data8	0.32	0.49	0.399	0.315	0.287	0.345	0.401	**0.496**
Data9	0.455	0.493	0.477	0.482	0.441	0.444	**0.497**	0.495
Data10	0.377	0.569	0.493	0.124	0.827	0.868	**0.872**	0.56
Data11	0.778	0.815	0.713	0.619	0.686	0.735	0.759	**0.817**
Data12	0.682	0.87	0.848	0.813	0.728	0.87	**0.922**	0.834
Average	0.639	0.703	0.667	0.558	0.649	0.683	**0.706**	0.703

9 datasets and average running time comparing with other algorithms. So, overall, FSPA-MOID outperforms other feature selection algorithms in terms of running time and classification performance.

The above are all the results of the experimental comparison. From all the results, FSPA-MOID has excellent performance in feature selection.

5. Conclusion

Multi-source ordered information decision systems, as prevalent forms of information systems in everyday applications, warrant increased scholarly focus. These systems offer a more nuanced and enriched representation compared to their single-source counterparts, as they encapsulate a greater breadth of information value.

For the theoretical aspect, this paper addresses the feature selection challenge within multi-source ordered information decision systems by elucidating the dominant domain relationships and the ensuing upper and lower approximations inherent to ordered information systems. We adopt the approximation degree as a metric for attribute discernibility and harness the preference relations within the PROMETHEE method to construct a preference index matrix among attributes, thereby making feature selection via decision-making method. And for the practical aspect, as mentioned in the introduction of this paper, multi-source information systems are becoming increasingly common in real life under the context of big data. The feature selection method proposed in this paper will have more opportunities to be applied in real life, and can effectively select weighty attributes in multi-source ordered decision information systems while considering all information sources to achieve role of feature selection.

In our approach, each source within the multi-source ordered information decision system yields a distinct preference index matrix. These matrices are then equally weighted and integrated to form a

Table 17
The run time(s) of different algorithms for all datasets.

Dataset	HAR	WD-HAR	HFS-IVO	FS-MS	FSPA-MODI
Data1	961.091	81 725.251	548.768	52.687	**16.954**
Data2	2349.73	83 345.681	165.488	20.113	**7.157**
Data3	4.644	9.973	**0.081**	0.504	0.112
Data4	214.874	443.8	2.385	8.895	**2.179**
Data5	753.451	16 285.402	35.97	22.23	**8.14**
Data6	2485.54	98 357.335	3836.354	240.196	**97.877**
Data7	24.479	79.257	1.7	3.269	**1.24**
Data8	31.12	791.71	**4.658**	8.387	7.515
Data9	123.49	3308.394	55.216	55.037	**39.227**
Data10	3373.022	261 818.306	**965.371**	1318.445	1383.148
Data11	4328.492	44 152.241	898.418	1394.789	**857.449**
Data12	3905.136	4687.692	439.966	775.949	**398.429**
Average	1546.255	49 583.753	571.198	325.041	**232.626**

comprehensive matrix that encapsulates the attribute set's ordering for the entire system. The contributions of this work are threefold: (1) We introduce a novel feature selection methodology tailored for multi-source ordered information decision-making systems, grounded in the PROMETHEE approach; (2) Our method fully considers the information encapsulated within each source of a multi-source ordered information decision-making system; (3) Experimental validation on multiple datasets substantiates the efficacy of our feature selection algorithm.

However, the current algorithm uniformly assigns equal weights to each source, neglecting potential disparities in source significance, which could necessitate differential weighting. Future work will explore a more objective and scientifically rigorous method for source weighting, aiming to refine and enhance our approach accordingly.

CRediT authorship contribution statement

Weihua Xu: Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Zishuo Yang:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation.

Declaration of competing interest

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62376229) and Natural Science Foundation of Chongqing, China (No. CSTB2023NSCQ-LZX0027)

Data availability

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

References

Ahmadi, S.A., Peivandizadeh, A., 2022. Sustainable portfolio optimization model using promethee ranking: A case study of palm oil buyer companies. Discrete Dyn. Nat. Soc. 2022 (1), 8935213.
Anaraki, J.R., Eftekhari, M., 2013. Rough set based feature selection: A review. In: The 5th Conference on Information and Knowledge Technology. IEEE, pp. 301–306.
Behzadian, M., Kazemzadeh, R.B., Albadvi, A., Aghdasi, M., 2010. PROMETHEE: A comprehensive literature review on methodologies and applications. Eur. J. Oper. Res. 200 (1), 198–215.
Bogdanovic, D., Nikolic, D., Ilic, I., 2012. Mining method selection by integrated AHP and PROMETHEE method. An. Acad. Bras. Ciênc. 84, 219–233.
Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., Lang, M., 2020. Benchmark for filter methods for feature selection in high-dimensional classification data. Comput. Statist. Data Anal. 143, 106839.
Brans, J.P., Vincke, P., Mareschal, B., 1986. How to select and how to rank projects: The PROMETHEE method. European J. Oper. Res. 24 (2), 228–238.
Che, X., Mi, J., Chen, D., 2018. Information fusion and numerical characterization of a multi-source information system. Knowl.-Based Syst. 145, 121–133.
Chen, R.C., Dewi, C., Huang, S.W., Caraka, R.E., 2020. Selecting critical features for data classification based on machine learning methods. J. Big Data 7 (1), 52, Journal of Big Data, 2020 7(1): 52.
Choukolaei, H.A., Mirani, S.E., Hejazi, M., Ghasemi, P., 2023. A GIS-based crisis management using fuzzy cognitive map**: PROMETHEE approach (a potential earthquake in Tehran). Soft Comput. 1–22.
Deng, J., Zhan, J., Wu, W.Z., 2022. A ranking method with a preference relation based on the PROMETHEE method in incomplete multi-scale information systems. Inform. Sci. 608, 1261–1282.
Ghasemi, P., Mehdiabadi, A., Spulbar, C., Birau, R., 2021. Ranking of sustainable medical tourism destinations in Iran: An integrated approach using fuzzy SWARA-PROMETHEE. Sustainability 13 (2), 683.
Ghasemi, P., Talebi Brijani, E., 2014. An integrated FAHP-PROMETHEE approach for selecting the best flexible manufacturing system. Eur. Online J. Nat. Soc. Sci. 3 (4), 1137.
Gul, M., Celik, E., Gumus, A.T., Guneri, A.F., 2018. A fuzzy logic based PROMETHEE method for material selection problems. Beni-Suef Univ. J. Basic Appl. Sci. 7 (1), 68–79.
Guo, D., Xu, W., Ding, W., Yao, Y., Wang, X., Pedrycz, W., Qian, Y., 2024. Concept-cognitive learning survey: Mining and fusing knowledge from data. Inf. Fusion 109, 102426.
Guo, D., Xu, W., Qian, Y., Ding, W., 2023. M-FCCL: Memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion. Inf. Fusion 100, 101962.
Hu, M., Tsang, E.C., Guo, Y., Chen, D., Xu, W., 2022. Attribute reduction based on overlap degree and k-nearest-neighbor rough sets in decision information systems. Inform. Sci. 584, 301–324.
Jain, P., Tiwari, A.K., Som, T., 2020. A fitting model based intuitionistic fuzzy rough feature selection. Eng. Appl. Artif. Intell. 89, 103421.
Jensen, R., 2008. Rough set-based feature selection: A review. In: Rough Computing: Theories, Technologies and Applications. pp. 70–107.
Jiang, J., Zhang, X., Yang, J., 2022. Double-quantitative feature selection using bidirectional three-level dependency measurements in divergence-based fuzzy rough sets. Eng. Appl. Artif. Intell. 115, 105226.
Ju, H., Fan, X., Ding, W., Huang, J., Pedrycz, W., Yang, X., 2024. Multi-association evidential feature selection and its application to identifying schizophrenia. Inform. Sci. 674, 120647.
Li, W., Zhou, H., Xu, W., Wang, X.Z., Pedrycz, W., 2022. Interval dominance-based feature selection for interval-valued ordered data. IEEE Trans. Neural Netw. Learn. Syst. 34 (10), 6898–6912.
Ma, X.A., Xu, H., Liu, Y., Zhang, J.Z., 2024. Class-specific feature selection using fuzzy information-theoretic metrics. Eng. Appl. Artif. Intell. 136, 109035.
Modrzejewski, M., 1993. Feature selection using rough sets theory. In: Machine Learning: ECML-93: European Conference on Machine Learning Vienna, Austria, April 5–7, 1993 Proceedings. vol. 6, Springer Berlin Heidelberg, pp. 213–226.
Pan, Y., Xu, W., Ran, Q., 2023. An incremental approach to feature selection using the weighted dominance-based neighborhood rough sets. Int. J. Mach. Learn. Cybern. 14 (4), 1217–1233.
Pawlak, Z., Pawlak, Z., 1998. Rough set theory and its applications to data analysis. Cybern. Syst. 29 (7), 661–688, Cybernetics Systems, 29(7) (1998)661–688.
Qian, W., Xiong, Y., Ding, W., Huang, J., Vong, C.M., 2024. Label correlations-based multi-label feature selection with label enhancement. Eng. Appl. Artif. Intell. 127, 107310, Engineering Applications of Artificial Intelligence 127 (2024): 107310.
Qiu, Z., Zhao, H., 2022. A fuzzy rough set approach to hierarchical feature selection based on Hausdorff distance. Appl. Intell. 52 (10), 11089–11102.
Rao, H., Shi, X., Rodrigue, A.K., Feng, J., Xia, Y., Elhoseny, M., Yuan, X., Gu, L., 2019. Feature selection based on artificial bee colony and gradient boosting decision tree. Appl. Soft Comput. 74, 634–642, Applied Soft Computing, 74 (2019)634–642.
Saadi, I., Farooq, B., Mustafa, A., Teller, J., Cools, M., 2018. An efficient hierarchical model for multi-source information fusion. Expert Syst. Appl. 110, 352–362.

Sang, B., Chen, H., Li, T., Xu, W., Yu, H., 2020. Incremental approaches for heterogeneous feature selection in dynamic ordered data. Inform. Sci. 541, 475–501.

Sang, B., Chen, H., Yang, L., Li, T., Xu, W., Luo, C., 2021. Feature selection for dynamic interval-valued ordered data based on fuzzy dominance neighborhood rough set. Knowl.-Based Syst. 227, 107223.

Shen, Y., Ma, X., Zhan, J., 2023. A two-stage adaptive consensus reaching model by virtue of three-way clustering for large-scale group decision making. Inform. Sci. 649, 119658.

Swiniarski, R.W., Skowron, A., 2003. Rough set methods in feature selection and recognition. Pattern Recognit. Lett. 24 (6), 833–849.

Xu, W., Pan, Y., Chen, X., Ding, W., Qian, Y., 2022a. A novel dynamic fusion approach using information entropy for interval-valued ordered datasets. IEEE Trans. Big Data 9 (3), 845–859.

Xu, W., Yang, Y., 2023. Matrix-based feature selection approach using conditional entropy for ordered data set with time-evolving features. Knowl.-Based Syst. 279, 110947.

Xu, W., Yu, J., 2017. A novel approach to information fusion in multi-source datasets: A granular computing viewpoint. Inf. Sci. 378, 410–423.

Xu, W., Yuan, K., Li, W., Ding, W., 2022b. An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution. IEEE Trans. Emerg. Top. Comput. Intell. 7 (1), 76–88.

Yuan, K., Miao, D., Pedrycz, W., Ding, W., Zhang, H., 2024. Ze-HFS: Zentropy-based uncertainty measure for heterogeneous feature selection and knowledge discovery. IEEE Trans. Knowl. Data Eng..

Yuan, K., Miao, D., Yao, Y., Zhang, H., Zhao, X., 2023. Feature selection using zentropy-based uncertainty measure. IEEE Trans. Fuzzy Syst..

Zhang, P., Li, T., Wang, G., Luo, C., Chen, H., Zhang, J., . . ., Yu, Z., 2021. Multi-source information fusion based on rough set theory: A review. Inf. Fusion 68, 85–117.

Zhang, P., Li, T., Wang, G., Wang, D., Lai, P., Zhang, F., 2023. A multi-source information fusion model for outlier detection. Inf. Fusion 93, 192–208.

Zhang, P., Li, T., Yuan, Z., Luo, C., Liu, K., Yang, X., 2022a. Heterogeneous feature selection based on neighborhood combination entropy. IEEE Trans. Neural Netw. Learn. Syst. 35 (3), 3514–3527.

Zhang, P., Li, T., Yuan, Z., Luo, C., Wang, G., Liu, J., Du, S., 2022b. A data-level fusion model for unsupervised attribute selection in multi-source homogeneous data. Inf. Fusion 80, 87–103.

Zhong, N., Dong, J., Ohsuga, S., 2001. Using rough sets with heuristics for feature selection. J. Intell. Inf. Syst. 16, 199–214.