Information Sciences 652 (2024) 119660

Contents lists available at ScienceDirect

Information Sciences

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

Semi-supervised feature selection based on fuzzy related family

Zhijun Guo [a], Yang Shen [a], Tian Yang [a,*], Yuan-Jiang Li [a,*], Yanfang Deng [a], Yuhua Qian [b]

[a] Hunan Provincial Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha, Hunan 410081, China

[b] Institute of Big Data Science and Industry, Shanxi University, Taiyuan, Shanxi 030006, China

ARTICLE INFO

Keywords:
Rough sets
Fuzzy sets
Feature selection
Covering rough set
Partially labeled data
Related family

ABSTRACT

Current machine learning algorithms encounter challenges such as missing labels and high dimensionality. Feature selection serves as an effective dimensionality reduction technique, enhancing the efficiency and accuracy of subsequent machine learning tasks by eliminating irrelevant and redundant features. Given the difficulty in obtaining fully labeled data, partially labeled data has become a crucial target for machine learning models to address. The related family is an efficient, rough set-based feature selection approach; however, it cannot be applied to semi-supervised learning tasks. Consequently, this paper introduces a semi-supervised feature selection method based on a fuzzy related family for partially labeled data. At first, the fuzzy label values of unlabeled samples are calculated based on fuzzy similarity relationships by establishing a novel fuzzy covering system. Subsequently, a fuzzy related family is constructed by a consistent fuzzy set. Then a semi-supervised feature selection algorithm, referred to as the Semi-supervised Fuzzy Related Family (SFRF), is developed using the established feature significance measurement. Compared to existing semi-supervised feature selection algorithms, SFRF considerably enhances feature selection efficiency while preserving classification accuracy. Specifically, the average reduction efficiency across twelve datasets increased by up to 109 times.

1. Introduction

With the rapid advancement of computer technology, data-driven machine learning methods have been extensively employed in various fields, such as finance, healthcare, and transportation. Effective utilization of data has always been a key challenge in the field of machine learning, and feature selection can aid in eliminating redundant attributes, improving computational efficiency, and ensuring the effective utilization of data information. Therefore, feature selection is a popular preprocessing step in machine learning. Currently, most of feature selection methods are primarily applied to supervised learning tasks. These methods typically form the final feature subset by selecting features with high relevance to labels, considering the variance of the feature, observing the learning curves of the model, and so on. Thereby improve the efficiency and performance of subsequent classification learning and data mining tasks, which usually necessitate datasets containing all labels. However, obtaining labels for data is often a challenging and time-consuming task, resulting in most real-world datasets being partially labeled. Thus, semi-supervised feature selection methods for handling partially labeled data have garnered widespread attention in recent years [1–4].

* Corresponding author.
E-mail addresses: guozhijun@hunnu.edu.cn (Z. Guo), math_yangtian@126.com (T. Yang), bugbuster@hunnu.edu.cn (Y.-J. Li), dengyanfang21@126.com (Y. Deng), jinchengqyh@126.com (Y. Qian).

https://doi.org/10.1016/j.ins.2023.119660
Received 7 March 2023; Received in revised form 31 July 2023; Accepted 4 September 2023

Available online 17 September 2023
0020-0255/

Semi-supervised feature selection methods typically involve two crucial steps: handling unlabeled data and evaluating feature significance. In the first step, there are two primary approaches. (1) One approach predicts the labels of unlabeled samples, converting the data into fully labeled data. For example, An et al. [5] propose a method called kNN-FRS based on fuzzy rough set with a relative measurement, introducing a novel measure of classification uncertainty. Shu et al. [6] propose a method that calculates the ϵ-neighborhood of unlabeled samples and expand it to obtain a larger neighborhood. Furthermore, Campagner et al. [7] presente a weakly supervised feature selection method to handle labels with multiple possible conditions. To better process weakly labeled data, they delve into the fundamental mathematical connections between rough set theory and belief function theory [8] [9]. Researchers have also proposed various methods for label prediction [10] [11] [12]. Although these methods can effectively handle partially labeled data, their efficiency is relatively lower due to the costs associated with label prediction. (2) The other approach ignores unlabeled samples and selects informative features based on existing labels. Qian et al. [13,14] and Yang et al. [15] define several local rough set models based on existing labels to propose fast local feature selection algorithms. Dai et al. [16] introduce the concept of discernibility pairs and dependency to measure the significance of labeled and unlabeled data. Pand et al. [17] combine the neighborhood discrimination index with the Laplacian score to handle labeled and unlabeled samples. In the second step, the focus is on evaluating the relevance between features or feature subsets and decision labels. Shu et al. [18] propose a multi-criteria measure method that combines dependency, information entropy, and information granulation to comprehensively evaluate features. Liu et al. [19] adhere to the principles of maximizing relevance and minimizing redundancy and utilize a forward sequential search strategy to gradually identify qualified features. Qian et al. [20] flexibly generate neighbors using the granular-ball-constrained neighbor strategy and obtain the confidence of labels for samples.

Rough set theory, a granular computing model proposed by Pawlak [21], forms upper and lower approximations without requiring prior knowledge to approximately describe data structures. Two feature selection frameworks, dependency degree and discernibility matrix, have been developed based on rough set theory and used for semi-supervised feature selection [18,20,22–25]. However, the classical rough set model can only be applied to categorical data. The fuzzy set, proposed by Zadeh [26], can effectively handle continuous data without discretization and information loss. Combining the advantages of both models, Dubois and Prade [27] introduce the fuzzy rough set, which can simultaneously describe fuzziness and roughness, to process continuous data without losing information. Based on fuzzy rough sets, many feature selection methods have been proposed [28–34]. Hu et al. [35] propose a feature selection algorithm based on multi- kernel fuzzy rough set for large-scale multimodality data. Sun et al. [33] introduce an innovative approach to feature selection within neighborhood decision systems. Huang et al. [34] presente a multigranulation fuzzy rough set model and design a noise-tolerant feature selection algorithm by means of the fuzzy β-neighborhood. Especially, researchers have explored semi-supervised feature selection methods based on fuzzy rough sets. Ma et al. [36] propose a semi-supervised rough fuzzy Laplacian feature mapping method that combines neighborhood fuzzy rough sets with Laplacian. Xing et al. [37] propose a weighted fuzzy rough sets-based multi-view tri-training model for partially labeled data.

The related family is an efficient feature selection method based on rough set theory, proposed by Yang et al. [15,38–40]. Building upon the related family, Lang et al. [41,42] investigate the mechanism of feature selection in dynamic covering decision systems with object and feature changes. However, the current related family method is only applicable to supervised learning tasks and faces difficulties when dealing with partially labeled data. Furthermore, although fuzzy rough set theory can effectively represent knowledge and extract valuable information from continuous data, the combination of fuzzy rough set theory and the related family method has been less explored. Therefore, it is crucial to develop new approaches that integrate the strengths of both theories to advance the field of semi-supervised feature selection for partially labeled data.

In light of these challenges, a semi-supervised feature selection method is proposed based on the fuzzy related family, which can effectively utilize existing labeled data. Firstly, fuzzy labels are assigned to each unlabeled sample based on the fuzzy similarity relationship, rather than a single label. Unlike adding a single label to each unlabeled sample, fuzzy labels reflect the degree of membership of the sample to various fuzzy decision classes, which more closely aligns with real-world situations. Secondly, the paper introduces the concept of a fuzzy related family to address partially labeled datasets for the first time. Thirdly, an efficient semi-supervised feature selection algorithm is proposed. Numerical experimental results demonstrate that, compared to existing semi-supervised feature selection methods, the average computational efficiency of feature selection can be improved by up to 109 times.

2. Background knowledge

In this section, the basic concepts of fuzzy covering rough set and partially labeled data are introduced.

The fuzzy set, a mathematical model proposed by Zadeh [26] in 1965, is designed to describe the fuzziness of data. Based on different membership functions, the inclusion relationship between a sample set and each sample in the universe is mapped onto the interval $[0,1]$, rather than $\{0,1\}$. Compared to traditional set theory, the fuzzy set can better describe uncertainty and fuzziness in the real world. To address the issue that classical rough set theory cannot handle continuous data, Dubois and Prade [27] introduce fuzzy set into rough set theory and propose the concept of fuzzy rough set. Covering is an important concept in rough set theory, as it represents the set of information granules. To better express continuous data, fuzzification of covering has emerged as a popular research topic in both fuzzy theory and rough set theory [43–45].

Definition 2.1. [43] Given a nonempty finite set U (called universe) and the fuzzy power set $\mathscr{F}(U)$ of U. For $\widetilde{FC}=\{\widetilde{F}_1,\widetilde{F}_2,\cdots,\widetilde{F}_s\}$, where fuzzy sets $\widetilde{F}_i\in\mathscr{F}(U)$ and $i=1,2,...,s$, if for any $w\in U$, $(\bigcup_{i=1}^{s}\widetilde{F}_i)(w)=1$, then $\widetilde{FC}$ is called a fuzzy covering of U.

In this paper, the union between the two fuzzy sets $\widetilde{F}_i \cup \widetilde{F}_j$ indicates that for any $w \in U$, there exist $(\widetilde{F}_i \cup \widetilde{F}_j)(w) = max\{\widetilde{F}_i(w), \widetilde{F}_j(w)\}$.

Ma et al. [44] point out that the condition $(\bigcup_{i=1}^{s} \widetilde{F}_i)(w) = 1$ is too strict, so they extend the fuzzy covering to the fuzzy β-covering, using the parameter β $(0 < \beta \leq 1)$ to replace 1. In this paper, we set $\beta = 0$ for the convenience of parameter choice and adjust the corresponding equation. The reason why β is set as 0 has been discussed in our published paper [46]. Thus, in order to allow β take the value of 0, we modified all occurrences of "$\geq \beta$" encountered in Definition 2.1 and Definition 2.2 to be "$> \beta$" in the article.

Definition 2.2. [44] Given a universe U and the fuzzy power set $\mathscr{F}(U)$ on U. For $\widetilde{FC} = \{\widetilde{F}_1, \widetilde{F}_2, \cdots, \widetilde{F}_s\}$, where $\widetilde{F}_i \in \mathscr{F}(U)$, $i = 1, 2, ..., s$, if for any $w \in U$, $(\bigcup_{i=1}^{s} \widetilde{F}_i)(w) > \beta$, where $\beta \in [0, 1]$, then $\widetilde{FC}$ is called a fuzzy β-covering of U, $(U, \widetilde{FC})$ is an approximate space of fuzzy β-covering.

Generally, describing an object (sample) only requires information related to this object (sample), not all information. So Yang et al. [45] propose the concept of fuzzy β-minimal description.

Definition 2.3. [45] Given an universe U and a fuzzy β-covering $\widetilde{FC} = \{\widetilde{F}_1, \widetilde{F}_2, \cdots, \widetilde{F}_s\}$ on U, where $\beta \in [0, 1]$. For any $w \in U$, the fuzzy β minimal description of w is defined as:

$$(\mathcal{M}_w^{\beta})_{\widetilde{FC}} = \{\widetilde{F} \in \widetilde{FC} | (\widetilde{F}(w) > \beta) \wedge (\forall \widetilde{H} \in \widetilde{FC} \wedge \widetilde{H}(w) > \beta \wedge \widetilde{H} \subseteq \widetilde{F} \Rightarrow \widetilde{F} = \widetilde{H})\} \tag{1}$$

The relationship $\widetilde{H} \subseteq \widetilde{F}$ between the two fuzzy sets denoted that for $\forall w \in U$, $\widetilde{H}(w) \leq \widetilde{F}(w)$. The subscript $\widetilde{FC}$ can be omitted without causing confusion.

Definition 2.4. Let $(U, \widetilde{FC})$ be an approximate space of fuzzy β-covering, where $\beta \in [0, 1]$. For any $\widetilde{W} \in \mathscr{F}(U)$, the fuzzy β-covering lower approximation $\underline{FL}_{\widetilde{FC}}(\widetilde{W})$ and the upper approximation $\overline{FH}_{\widetilde{FC}}(\widetilde{W})$ of $\widetilde{W}$ are defined as:

$$\underline{FL}_{\widetilde{FC}}(\widetilde{W}) = \cup\{\widetilde{F} \in \widetilde{FC} | \widetilde{F} \subseteq \widetilde{W}\} \tag{2}$$

$$\overline{FH}_{\widetilde{FC}}(\widetilde{W}) = \underline{FL}_{\widetilde{FC}}(\widetilde{W}) \cup (\cup\{\mathcal{M}_w^{\beta} | \widetilde{W}(w) > \beta\}) \tag{3}$$

In this paper, the fuzzy β-covering lower approximation is defined as Definition 2.4. For convenience, we set β as 0.

3. A new semi-supervised feature selection method

To fully utilize the potential information of unlabeled samples, this chapter presents a semi-supervised feature selection method based on fuzzy related families. Firstly, a partially labeled dataset is converted into a fuzzy label covering information system by evaluating the fuzzy similarity between labeled samples and unlabeled samples. Next, a fuzzy related family is constructed based on the fuzzy label covering information system. As a result, a feature evaluation function is defined to select informative features.

3.1. Fuzzy label

In cases involving partially labeled data, predicting the missing labels can be achieved by utilizing the available labeled data. More precisely, if an unlabeled sample exhibits similarity to certain labeled samples, it is probable that its label will also resemble those of the labeled samples. This relationship of similarity between samples can offer valuable information about the underlying patterns in the data. Consequently, the fuzzy similarity relation is employed to characterize this similarity relationship, serving not only to indicate whether two samples are similar but also to quantify the degree of similarity. In this section, we explore adding fuzzy labels for unlabeled samples based on fuzzy similarity relations.

Let $(U^l \cup U^u, AT, D)$ be a partial label information system, where U^l, U^u, AT and D are the set of labeled samples, the set of unlabeled samples, the conditional feature set and the decision feature, respectively. FN_{AT} is a fuzzy similarity relation based on the conditional feature set AT. For $\forall w_i, w_j \in U^l \cup U^u$, FN_{AT} satisfies: (1) reflexivity: $FN_{AT}(w_i, w_i) = 1$; (2) symmetry: $FN_{AT}(w_i, w_j) = FN_{AT}(w_j, w_i)$. In this paper, the subscript AT is omitted without causing confusion.

For $\forall w_i, w_j \in U^l \cup U^u$, the fuzzy similarity between sample w_i and sample w_j is calculated by:

$$FN(w_i, w_j) = \begin{cases} 1 - df(w_i, w_j), & df(w_i, w_j) < 1 \\ 0, & otherwise \end{cases} \tag{4}$$

where $df(w_i, w_j) = (\sum_{a \in AT} |a(w_i) - a(w_j)|^2)^{1/2}$ is the Euclidean distance function, and the values under each feature have been normalized to $[0, 1]$. The Euclidean distance is applied as an example, and other metric distances or kernels are also viable options.

Suppose the set of labeled samples U^l can be divided into s categories by the labeled features D, i.e., $U^l/D = \{D_1, D_2, \cdots, D_s\}$. Then the corresponding fuzzy set $\widetilde{D}_i$ is obtained by converting the classical sets into fuzzy. For labeled samples, if $w \in D_i$, since w has a label, then $\widetilde{D}_i(w) = 1$, and $\widetilde{D}_j(w) = 0$ for any $j \neq i$. For an unlabeled sample $\forall w \in U^u$, let k be the parameter of the number of nearest labeled samples, k is set as 10% of the number of labeled samples in the experiments. Let $N_w = \{z_1, z_2, \cdots, z_k\} \subseteq U^l$ be the set of k nearest samples, then calculate the membership degree $\widetilde{D}_i(w)$ that w belonging to the category D_i based on the fuzzy similarity between w and samples in N_w. The detailed description is shown in Definition 3.1.

Table 1
Partial Labeled Data.

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}
a_1	0.2763	0.2552	0.4868	0.8605	0.3657	0.8816	0.4578	0.2763	0.2973	0.7500
a_2	0.2154	0.5316	0.4446	0.2332	0.3577	0.2233	0.3260	0.2648	0.1719	0.8498
a_3	0.5133	0.3422	0.5561	0.7272	0.4866	0.5454	0.4919	0.1818	0.5080	0.4652
a_4	0.4072	0.4329	0.4845	0.4845	0.5876	0.0721	0.4587	0.3556	0.6288	0.4845
d	2	2	3	1	2	1	3	2	2	3

	w_{11}	w_{12}	w_{13}	w_{14}	w_{15}	w_{16}	w_{17}	w_{18}	w_{19}	w_{20}
a_1	0.8157	0.7473	0.7500	0.8158	0.7474	0.3526	0.6868	0.3211	0.3921	0.6368
a_2	0.6640	0.2290	0.8498	0.6640	0.2292	0.0395	0.4664	0.1957	0.3340	0.5850
a_3	0.7379	0.7700	0.4652	0.7380	0.7701	0.0000	0.6417	0.4064	0.4332	0.6631
a_4	0.7164	0.4536	0.4845	0.7165	0.4536	0.0000	0.2371	0.4330	0.5361	0.6392
d	3	1	*	*	*	*	*	*	*	*

Table 2
Fuzzy Membership Table of Decision Class.

	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}
$\widetilde{D}_1$	0	0	0	1	0	1	0	0	0	0
$\widetilde{D}_2$	1	1	0	0	1	0	0	1	1	0
$\widetilde{D}_3$	0	0	1	0	0	0	1	0	0	1

	w_{11}	w_{12}	w_{13}	w_{14}	w_{15}	w_{16}	w_{17}	w_{18}	w_{19}	w_{20}
$\widetilde{D}_1$	0	1	0.1994	0.2181	0.4166	0.1053	0.3413	0.1761	0.1357	0.1949
$\widetilde{D}_2$	0	0	0.2233	0.2362	0.2415	0.6987	0.2521	0.5967	0.5639	0.3332
$\widetilde{D}_3$	1	0	0.5773	0.5458	0.3419	0.1960	0.4066	0.2272	0.3004	0.4719

Definition 3.1. Given a partial label information system $(U^l \cup U^u, AT, D)$ and the division of the labeled sample sets on the label features $U^l/D = \{D_1, D_2, \cdots, D_s\}$. For $\forall w \in U^l \cup U^u$, $D_i \in U^l/D$, the fuzzy label value $\widetilde{D}_i(w)$ of sample w is defined as:

$$\widetilde{D}_i(w) = \begin{cases} \frac{\sum_{z \in N_w \cap D_i} FN(w,z)}{\sum_{z \in N_w} FN(w,z)}, & if\ \ w \in U^u \\ 1, & if\ \ w \in U^l\ \ and\ \ w \in D_i \\ 0, & if\ \ w \in U^l\ \ and\ \ w \notin D_i \end{cases} \tag{5}$$

where FN is a fuzzy similarity relation based on the conditional feature set AT, $N_w = \{z_1, z_2, \cdots, z_k\}$ are the k closest labeled samples to w. The fuzzy labels of the universe $U^l \cup U^u$ are then denoted as: $\widetilde{FD} = \{\widetilde{D}_1, \widetilde{D}_2, \cdots, \widetilde{D}_s\}$.

Example 3.1. Given a partial label information system $PLIT = (U^l \cup U^u, AT, D)$, where the labeled sample set $U^l = \{w_1, w_2, \cdots, w_{12}\}$, the unlabeled sample set $U^u = \{w_{13}, w_{14}, \cdots, w_{20}\}$, the condition features $AT = \{a_1, a_2, a_3, a_4\}$, and the decision feature $D = \{d\}$. The details are shown in Table 1. It is easy to get $U^l/D = \{D_1, D_2, D_3\}$: $D_1 = \{w_4, w_6, w_{12}\}$, $D_2 = \{w_1, w_2, w_5, w_8, w_9\}$, $D_3 = \{w_3, w_7, w_{10}, w_{11}\}$.

For the labeled sample w_1, since $w_1 \in D_2$, we get $\widetilde{D}_2(w_1) = 1$ and $\widetilde{D}_1(w_1) = \widetilde{D}_3(w_1) = 0$; similarly, we can obtain the fuzzy labels of all labeled samples.

For the unlabeled sample w_{13}, if $k = 10$, the Euclidean distance function is used to find the 10 closest samples to w_{13} in the set of labeled samples U^l to obtain $N_{w_{12}} = \{w_{10}, w_{11}, w_3, w_7, w_2, w_5, w_4, w_{12}, w_6, w_1\}$ (if the number of labeled samples is less than 10, then all labeled samples are selected). Then the fuzzy similarity between these 10 samples and sample w_{13} is calculated respectively: $FN(w_{13}, w_{10}) = 1$, $FN(w_{13}, w_{11}) = 0.5914$, $FN(w_{13}, w_3) = 0.5084$, $FN(w_{13}, w_7) = 0.3991$, and $FN(w_{13}, w_2) = 0.3968$, $FN(w_{13}, w_5) = 0.3668$, $FN(w_{13}, w_4) = 0.321$, $FN(w_{13}, w_{12}) = 0.3079$, $FN(w_{13}, w_{12}) = 0.3079$, $FN(w_{13}, w_{12}) = 0.3079$, $FN(w_{13}, w_{12}) = 0.3079$. $FN(w_{13}, w_6) = 0.2343$, $FN(w_{13}, w_1) = 0.203$, and $FN(w_{13}, w_1) = 0.203$. Then the fuzzy labels of the sample w_{13} are calculated according to the Equation (5) as $\widetilde{D}_1(w_{13}) = 0.1994$, $\widetilde{D}_2(w_{13}) = 0.2233$, $\widetilde{D}_3(w_{13}) = 0.5773$.

Similarly, the fuzzy label values can be derived for all unlabeled samples with respect to all labeled categories. The final result is shown in Table 2.

3.2. Fuzzy label covering information system

Given a partial label information system $(U^l \cup U^u, AT, D)$, the conditional feature $AT = \{a_1, a_2, \cdots, a_n\}$. Based on each conditional feature $a_i \in AT$, we can form the corresponding fuzzy covering $\widetilde{FC}_i$, then these fuzzy coverings are used to form a fuzzy covering family $\mathscr{V} = \{\widetilde{FC}_1, \widetilde{FC}_2, \cdots, \widetilde{FC}_n\}$ on universe $U^l \cup U^u$. Then all fuzzy labels $\widetilde{FD} = \{\widetilde{D}_1, \widetilde{D}_2, \cdots, \widetilde{D}_s\}$ of universe $U^l \cup U^u$ are calculated based on Definition 3.1. Thus we can convert $(U^l \cup U^u, AT, D)$ to $(U^l \cup U^u, \mathscr{V}, \widetilde{FD})$, called FLCIS (Fuzzy Labeled Covering Information System).

Suppose that $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is a FLCIS, $\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \cdots, \widetilde{\mathcal{FC}}_n\}$ is a fuzzy covering family of $U^l \cup U^u$. Let $\cup\mathscr{V} = \{\widetilde{F} \in \widetilde{\mathcal{FC}}_i | \widetilde{\mathcal{FC}}_i \in \mathscr{V}\}$, because $\widetilde{\mathcal{FC}}_i \in \mathscr{V}$ is a fuzzy covering on universe $U^l \cup U^u$, $\cup\mathscr{V}$ is also a covering on universe $U^l \cup U^u$. For any fuzzy set $\widetilde{W} \in \mathscr{F}(U^l \cup U^u)$, the fuzzy covering lower approximation of $\widetilde{W}$ about $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is defined as:

$$\underline{FL}_{\cup\mathscr{V}}(\widetilde{W}) = \cup\{\widetilde{F} \in \cup\mathscr{V} | \widetilde{F} \subseteq \widetilde{W}\} \tag{6}$$

Based on the fuzzy inclusion operation, the fuzzy positive region of fuzzy label $\widetilde{\mathcal{FD}}$ is defined as:

$$POS_{\cup\mathscr{V}}(\widetilde{\mathcal{FD}}) = \cup\{\underline{FL}_{\cup\mathscr{V}}(\widetilde{D}_i) | \widetilde{D}_i \in \widetilde{\mathcal{FD}}\} = \cup\{\widetilde{F} \in \cup\mathscr{V} | \exists \widetilde{D}_i \in \widetilde{\mathcal{FD}} \text{ s.t. } \widetilde{F} \subseteq \widetilde{D}_i\} \tag{7}$$

Generally, the objective of feature selection is to identify the minimal subset that maintains the positive region invariant. Given that retaining the fuzzy positive region of fuzzy covering entirely invariant is overly stringent, the coverage rate regarding the fuzzy positive region is proposed as a constraint for feature selection. In this study, the coverage rate concerning the fuzzy positive region is defined within the context of the fuzzy label covering information system.

Definition 3.2. Given a FLCIS $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, where $\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \cdots, \widetilde{\mathcal{FC}}_n\}$ is a fuzzy covering family on $U^l \cup U^u$, $\widetilde{\mathcal{FD}} = \{\widetilde{D}_1, \widetilde{D}_2, \cdots, \widetilde{D}_s\}$ are fuzzy labels of $U^l \cup U^u$. The fuzzy positive region coverage set $CS(\mathscr{T})$ and the fuzzy positive region coverage rate $CR(\mathscr{T})$ for any $\mathscr{T} \subseteq \mathscr{V}$ are:

$$CS(\mathscr{T}) = \{w_i | POS_{\cup\mathscr{T}}(\widetilde{\mathcal{FD}})(w_i) > 0\} \tag{8}$$

$$CR(\mathscr{T}) = |CS(\mathscr{T})| / |U^l \cup U^u| \tag{9}$$

Proposition 3.1. *Let $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ be a FLCIS, where $\mathscr{V}$ is a fuzzy covering family of $U^l \cup U^u$. If $\mathscr{T}_1 \subseteq \mathscr{T}_2 \subseteq \mathscr{V}$, then $CR(\mathscr{T}_1) \le CR(\mathscr{T}_2)$.*

Proof. Since $POS_{\cup\mathscr{T}_2}(\widetilde{\mathcal{FD}}) = \cup\{\underline{FL}_{\cup\mathscr{T}_2}(\widetilde{D}_i) | \widetilde{D}_i \in \widetilde{\mathcal{FD}}\} = \cup\{\widetilde{F} \in \cup\mathscr{T}_2 | \exists \widetilde{D}_i \in \widetilde{\mathcal{FD}} \text{ s.t. } \widetilde{F} \subseteq \widetilde{D}_i\} = (\cup\{\widetilde{F} \in \cup\mathscr{T}_1 | \exists \widetilde{D}_i \in \widetilde{\mathcal{FD}} \text{ s.t. } \widetilde{F} \subseteq \widetilde{D}_i\}) \cup (\cup\{\widetilde{F} \in \cup(\mathscr{T}_2 - \mathscr{T}_1) | \exists \widetilde{D}_i \in \widetilde{\mathcal{FD}} \text{ s.t. } \widetilde{F} \subseteq \widetilde{D}_i\})$. It is obvious that $POS_{\cup\mathscr{T}_2}(\widetilde{\mathcal{FD}}) = POS_{\cup\mathscr{T}_1}(\widetilde{\mathcal{FD}}) \cup (\cup\{\widetilde{F} \in \cup(\mathscr{T}_2 - \mathscr{T}_1) | \exists \widetilde{D}_i \in \widetilde{\mathcal{FD}} \text{ s.t. } \widetilde{F} \subseteq \widetilde{D}_i\})$, i.e. $POS_{\cup\mathscr{T}_1}(\widetilde{\mathcal{FD}}) \subseteq POS_{\cup\mathscr{T}_2}(\widetilde{\mathcal{FD}})$. Thus $CS(\mathscr{T}_1) \subseteq CS(\mathscr{T}_2)$, then $CR(\mathscr{T}_1) \le CR(\mathscr{T}_2)$, so the fuzzy positive region coverage rate is monotonic. □

Definition 3.3. Given a FLCIS $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, where $\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \cdots, \widetilde{\mathcal{FC}}_n\}$ is a fuzzy covering family on $U^l \cup U^u$. For any $\widetilde{\mathcal{FC}}_i \in \mathscr{V}$, if $CR(\mathscr{V} - \{\widetilde{\mathcal{FC}}_i\}) = CR(\mathscr{V})$, $\widetilde{\mathcal{FC}}_i$ is said to be reducible in $\mathscr{V}$, otherwise $\widetilde{\mathcal{FC}}_i$ is necessary in $\mathscr{V}$. And for any $\mathscr{T} \subseteq \mathscr{V}$, if every element in the $\mathscr{T}$ is necessary, then $\mathscr{T}$ is independent. If $\mathscr{T}$ is independent, and $CR(\mathscr{T}) = CR(\mathscr{V})$, then $\mathscr{T}$ is said to be a reduct of $\mathscr{V}$.

For any $\mathscr{T} \subseteq \mathscr{V}$, it is evident that $CS(\mathscr{T}) \subseteq CS(\mathscr{V})$. If $CS(\mathscr{V}) \subseteq CS(\mathscr{T})$, it is clear that $CS(\mathscr{T}) = CS(\mathscr{V})$. Therefore, the purpose of feature selection in this paper is to find a minimal feature subset (reduct) keeping $CS(\mathscr{V})$ or $CR(\mathscr{V})$ invariant.

3.3. Fuzzy related family

The related family, an efficient feature selection method proposed by Yang et al. [38], is based on covering rough sets. However, this method is unable to process partially labeled data. To address this issue, we propose a novel approach called the fuzzy related family, which is based on a fuzzy label covering information system. This approach effectively designs a feature evaluation function for selecting relevant features.

Definition 3.4. Given a FLCIS $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, where $\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \cdots, \widetilde{\mathcal{FC}}_n\}$, is a family of fuzzy coverings on $U^l \cup U^u$. For any sample $w_i \in U^l \cup U^u$, define the related set of w_i as $r(w_i) = \{\widetilde{\mathcal{FC}} \in \mathscr{V} | \exists \widetilde{F} \in \widetilde{\mathcal{FC}} \wedge \exists \widetilde{D}_j \in \widetilde{\mathcal{FD}} \text{ s.t. } \widetilde{F} \subseteq \widetilde{D}_j \wedge \widetilde{F}(w_i) > 0\}$, the related family of fuzzy label covering information systems $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is $\mathbb{R}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}}) = \{r(w_i) | w_i \in U^l \cup U^u\}$.

The fuzzy related family introduces fuzzy concepts to represent the likelihood of unknown samples belonging to some consistent granules (which means unknown samples may be classified to the correct class), while the original related family just denote whether unknown samples belong to some consistent granules. In other words, when there are multiple features that could potentially classify a sample correctly, the fuzzy related family can assess which feature is more likely to lead to a correct classification, but the original related family choose one feature randomly.

Suppose a fuzzy information granule $\widetilde{F}$ and a fuzzy label $\widetilde{D}_j \in \widetilde{\mathcal{FD}}$, if $\widetilde{F} \subseteq \widetilde{D}_j$, we regard that the fuzzy information granule $\widetilde{F}$ is consistent to the fuzzy label $\widetilde{D}_j$ in this paper. As shown in Definition 3.4, the related set of a sample w_i consists of all feature generating at least one fuzzy information granules $\widetilde{F}$ which is consistent to a certain fuzzy label $\widetilde{D}_j$ and $\widetilde{F}(w_i) > 0$.

The related family of a fuzzy label covering information system $\mathbb{R}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}}) = \{r(w_i) | w_i \in U^l \cup U^u\}$ stores useful features for each sample, where $r(w_i)$ is the set of all feature classifying w_i correctly. If a feature is not included in a related set of any sample, it is totally useless for the fuzzy information system. However, the related family cannot describe the credibility degree of the correct classification. To solve this problem, the fuzzy related family $\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is defined to select informative feature more credible.

Definition 3.5. Given a FLCIS $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, where $\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \cdots, \widetilde{\mathcal{FC}}_n\}$ is a family of fuzzy coverings of $U^l \cup U^u$. For any $\widetilde{\mathcal{FC}}_i \in \mathscr{V}$, the consistent fuzzy set $\mathcal{R}(\widetilde{\mathcal{FC}}_i)$ of $\widetilde{\mathcal{FC}}_i$ is defined as:

$$\mathcal{R}(\widetilde{\mathcal{FC}}_i)(w_t) = \begin{cases} max\{\widetilde{F}(w_t) | \widetilde{F} \in \widetilde{\mathcal{FC}}_i \wedge \exists \widetilde{D}_j \in \widetilde{\mathcal{FD}} \text{ s.t. } \widetilde{F} \subseteq \widetilde{D}_j\}, & \widetilde{\mathcal{FC}}_i \in r(w_t) \\ 0, & otherwise \end{cases} \tag{10}$$

where w_t is a sample, $r(w_t) \in \mathbb{R}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, and $\mathbb{R}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is the related family of fuzzy label covering information systems $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$. Equation (11) can also be used to represent the consistent fuzzy set of $\widetilde{\mathcal{FC}}_i$:

$$\mathcal{R}(\widetilde{\mathcal{FC}}_i) = \cup\{\widetilde{F} \in \widetilde{\mathcal{FC}}_i | \exists \widetilde{D}_j \in \widetilde{\mathcal{FD}} \text{ s.t. } \widetilde{F} \subseteq \widetilde{D}_j\} \tag{11}$$

the fuzzy related family $\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is defined as:

$$\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}}) = \{\mathcal{R}(\widetilde{\mathcal{FC}}_i) | \widetilde{\mathcal{FC}}_i \in \mathscr{V}\} \tag{12}$$

The related family $\mathbb{R}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ can be generated by the fuzzy related family $\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ over the following method shown in Proposition 3.2, which means the fuzzy related family contains more information about the data than the non-fuzzy one.

Proposition 3.2. *Given a FLCIS $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, $\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is the fuzzy related family. For each $w_t \in U^l \cup U^u$, the related set $r(w_t)$ can be induced by the fuzzy related family: $r(w_t) = \{\widetilde{\mathcal{FC}}_i | \mathcal{R}(\widetilde{\mathcal{FC}}_i) \in \mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}}), \mathcal{R}(\widetilde{\mathcal{FC}}_i)(w_t) > 0\}$. Then, the related family $\mathbb{R}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}}) = \{r(w_i) | w_i \in U^l \cup U^u\}$ can be generated by $\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$.*

Example 3.2. Given a FLCIS $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, where $U^l = \{w_1, w_2, w_3\}$, $U^u = \{w_4, w_5, w_6\}$, a fuzzy covering family $\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \widetilde{\mathcal{FC}}_3, \widetilde{\mathcal{FC}}_4\}$, $\widetilde{\mathcal{FC}}_1 = \{\widetilde{F}_{11}, \widetilde{F}_{12}, \widetilde{F}_{13}\}$, $\widetilde{\mathcal{FC}}_2 = \{\widetilde{F}_{21}, \widetilde{F}_{22}, \widetilde{F}_{23}\}$, $\widetilde{\mathcal{FC}}_3 = \{\widetilde{F}_{31}, \widetilde{F}_{32}, \widetilde{F}_{33}\}$, $\widetilde{\mathcal{FC}}_4 = \{\widetilde{F}_{41}, \widetilde{F}_{42}, \widetilde{F}_{43}\}$ and fuzzy label $\widetilde{\mathcal{FD}} = \{\widetilde{D}_1, \widetilde{D}_2\}$. All fuzzy granules in each fuzzy covering are listed as

$\widetilde{F}_{11} = \frac{0.5}{w_1} + \frac{0.8}{w_2} + \frac{0.6}{w_3} + \frac{0.7}{w_4} + \frac{0.3}{w_5} + \frac{0.6}{w_6}$, $\widetilde{F}_{12} = \frac{0}{w_1} + \frac{0}{w_2} + \frac{0.9}{w_3} + \frac{0}{w_4} + \frac{0.7}{w_5} + \frac{0}{w_6}$,
$\widetilde{F}_{13} = \frac{0.8}{w_1} + \frac{0.6}{w_2} + \frac{0}{w_3} + \frac{0.8}{w_4} + \frac{0.1}{w_5} + \frac{0}{w_6}$, $\widetilde{F}_{21} = \frac{0.3}{w_1} + \frac{0.6}{w_2} + \frac{0}{w_3} + \frac{0.5}{w_4} + \frac{0.2}{w_5} + \frac{0}{w_6}$,
$\widetilde{F}_{22} = \frac{0}{w_1} + \frac{0.7}{w_2} + \frac{0.8}{w_3} + \frac{0.5}{w_4} + \frac{0.6}{w_5} + \frac{0}{w_7}$, $\widetilde{F}_{23} = \frac{0}{w_1} + \frac{0}{w_2} + \frac{0.2}{w_3} + \frac{0.2}{w_4} + \frac{0.6}{w_5} + \frac{0}{w_6}$,
$\widetilde{F}_{31} = \frac{0.4}{w_1} + \frac{0.6}{w_2} + \frac{0}{w_3} + \frac{0.7}{w_4} + \frac{0.8}{w_5} + \frac{0.7}{w_6}$, $\widetilde{F}_{32} = \frac{0}{w_1} + \frac{0}{w_2} + \frac{0.6}{w_3} + \frac{0}{w_4} + \frac{0.7}{w_5} + \frac{0.3}{w_6}$,
$\widetilde{F}_{33} = \frac{0.7}{w_1} + \frac{0.6}{w_2} + \frac{0}{w_3} + \frac{0}{w_4} + \frac{0.3}{w_5} + \frac{0.6}{w_6}$, $\widetilde{F}_{41} = \frac{0.5}{w_1} + \frac{0.5}{w_2} + \frac{0}{w_3} + \frac{0.6}{w_4} + \frac{0.2}{w_5} + \frac{0.5}{w_6}$,
$\widetilde{F}_{42} = \frac{0.3}{w_1} + \frac{0.7}{w_2} + \frac{0.6}{w_3} + \frac{0.1}{w_4} + \frac{0}{w_5} + \frac{0.3}{w_6}$, $\widetilde{F}_{43} = \frac{0.6}{w_1} + \frac{0}{w_2} + \frac{0.5}{w_3} + \frac{0.6}{w_4} + \frac{0.6}{w_5} + \frac{0.7}{w_6}$.

The fuzzy labels are

$\widetilde{D}_1 = \frac{1}{w_1} + \frac{1}{w_2} + \frac{0}{w_3} + \frac{0.8}{w_4} + \frac{0.3}{w_5} + \frac{0.6}{w_6}$, $\widetilde{D}_2 = \frac{0}{w_1} + \frac{0}{w_2} + \frac{1}{w_3} + \frac{0.2}{w_4} + \frac{0.7}{w_5} + \frac{0.4}{w_6}$.

The fuzzy positive region is calculated from the Equation (7) as follows,

$POS_{\cup\mathscr{V}}(\widetilde{\mathcal{FD}}) = \widetilde{F}_{12} \cup \widetilde{F}_{13} \cup \widetilde{F}_{21} \cup \widetilde{F}_{23} \cup \widetilde{F}_{32} \cup \widetilde{F}_{33} \cup \widetilde{F}_{41} = \frac{0.8}{w_1} + \frac{0.6}{w_2} + \frac{0.9}{w_3} + \frac{0.8}{w_4} + \frac{0.7}{w_5} + \frac{0.6}{w_6}$.

Find the related set for each $w_i \in U^l \cup U^u$ according to the Definition 3.4.

$r(w_1) = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \widetilde{\mathcal{FC}}_3, \widetilde{\mathcal{FC}}_4\}$, $r(w_2) = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \widetilde{\mathcal{FC}}_3, \widetilde{\mathcal{FC}}_4\}$,
$r(w_3) = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \widetilde{\mathcal{FC}}_3\}$, $r(w_4) = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \widetilde{\mathcal{FC}}_4\}$,
$r(w_5) = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \widetilde{\mathcal{FC}}_3, \widetilde{\mathcal{FC}}_4\}$, $r(w_6) = \{\widetilde{\mathcal{FC}}_3, \widetilde{\mathcal{FC}}_4\}$.

Find the consistent fuzzy set for each $\widetilde{\mathcal{FC}}_i \in \mathscr{V}$ according to the Definition 3.5.

$\mathcal{R}(\widetilde{\mathcal{FC}}_1) = \widetilde{F}_{12} \cup \widetilde{F}_{13} = \frac{0.8}{w_1} + \frac{0.6}{w_2} + \frac{0.9}{w_3} + \frac{0.8}{w_4} + \frac{0.7}{w_5} + \frac{0}{w_6}$,
$\mathcal{R}(\widetilde{\mathcal{FC}}_2) = \widetilde{F}_{21} \cup \widetilde{F}_{23} = \frac{0.3}{w_1} + \frac{0.6}{w_2} + \frac{0.2}{w_3} + \frac{0.5}{w_4} + \frac{0.6}{w_5} + \frac{0}{w_6}$,
$\mathcal{R}(\widetilde{\mathcal{FC}}_3) = \widetilde{F}_{32} \cup \widetilde{F}_{33} = \frac{0.7}{w_1} + \frac{0.6}{w_2} + \frac{0.6}{w_3} + \frac{0}{w_4} + \frac{0.7}{w_5} + \frac{0.6}{w_6}$,
$\mathcal{R}(\widetilde{\mathcal{FC}}_4) = \widetilde{F}_{41} = \frac{0.5}{w_1} + \frac{0.5}{w_2} + \frac{0}{w_3} + \frac{0.6}{w_4} + \frac{0.2}{w_5} + \frac{0.5}{w_6}$.

Finally get $\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}}) = \{\mathcal{R}(\widetilde{\mathcal{FC}}_1), \mathcal{R}(\widetilde{\mathcal{FC}}_2), \mathcal{R}(\widetilde{\mathcal{FC}}_3), \mathcal{R}(\widetilde{\mathcal{FC}}_4)\}$.

Theorem 3.1. *Let $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ be a FLCIS, $\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is the corresponding fuzzy related family. For $\forall w_t \in U^l \cup U^u$, $w_t \in CS(\mathscr{V})$ if and only if $\exists \widetilde{\mathcal{FC}}_i \in \mathscr{V}$ makes $\mathcal{R}(\widetilde{\mathcal{FC}}_i)(w_t) > 0$.*

Proof. ($\Rightarrow$) Since $\exists \widetilde{\mathcal{FC}}_i \in \mathscr{V}$ makes $\mathcal{R}(\widetilde{\mathcal{FC}}_i)(w_t) > 0$, then there exists $\widetilde{F}_j \in \widetilde{\mathcal{FC}}_i$ makes $\widetilde{F}_j(w_t) > 0$, and $\exists \widetilde{D}_r \in \widetilde{\mathcal{FD}}$ makes $\widetilde{F}_j \subseteq \widetilde{D}_r$. Since $POS_{\cup\mathscr{V}}(\widetilde{\mathcal{FD}}) = \cup\{\widetilde{F} \in \cup\mathscr{V} | \exists \widetilde{D}_r \in \widetilde{\mathcal{FD}} \text{ s.t. } \widetilde{F} \subseteq \widetilde{D}_r\}$, and $\widetilde{F}_j \in \cup\mathscr{V}$, there is $POS_{\cup\mathscr{V}}(\widetilde{\mathcal{FD}})(w_t) > 0$, which implies $w_t \in CS(\mathscr{V})$.

($\Leftarrow$) Assume $w_t \in CS(\mathscr{V})$, then $POS_{\cup\mathscr{V}}(\widetilde{\mathcal{FD}})(w_t) > 0$. Thus, $\exists \widetilde{F}_j \in \cup\mathscr{V}$ makes $\widetilde{F}_j(w_t) > 0$, and $\exists \widetilde{D}_r \in \widetilde{\mathcal{FD}}$ makes $\widetilde{F}_j \subseteq \widetilde{D}_r$. This means that $\exists \widetilde{\mathcal{FC}}_i \in \mathscr{V}$ makes $\mathcal{R}(\widetilde{\mathcal{FC}}_i)(w_t) > 0$. $\square$

Theorem 3.1 demonstrates that it is easy to determine whether an object belongs to the coverage set of a feature by the fuzzy related family. Because the fuzzy related family contains all information to compute the coverage of feature subsets, enabling the derivation of a new reduction approach based on the fuzzy related family.

Theorem 3.2. *Let $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ be a FLCIS, $\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is the corresponding fuzzy related family. For $\mathscr{T} \subseteq \mathscr{V}$, $\mathscr{T}$ is a reduct of $\mathscr{V}$ if and only if $\mathscr{T}$ is a minimal subset satisfying the condition: for any nonempty related set $r(w_t) \in \mathbb{R}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, there $\exists \widetilde{\mathcal{FC}}_r \in \mathscr{T}$ such that $\mathcal{R}(\widetilde{\mathcal{FC}}_r)(w_t) > 0$.*

Proof. ($\Rightarrow$) Assume that $\mathscr{T}$ is a reduct of $\mathscr{V}$, then $CS(\mathscr{T}) = CS(\mathscr{V})$ and $CR(\mathscr{T}) = CR(\mathscr{V})$. For $\forall r(w_t) \in \mathbb{R}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ and $r(w_t)$ is nonempty, then $w_t \in CS(\mathscr{V})$. And since $CS(\mathscr{T}) = CS(\mathscr{V})$, $w_t \in CS(\mathscr{T})$, which means that $\exists \widetilde{\mathcal{FC}}_r \in \mathscr{T}$ such that $\mathcal{R}(\widetilde{\mathcal{FC}}_r)(w_t) > 0$.

($\Leftarrow$) Suppose that for $\forall r(w_t) \in \mathbb{R}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, $r(w_t)$ is nonempty, there $\exists \widetilde{\mathcal{FC}}_r \in \mathscr{T}$ and $r(w_t) \in \mathbb{R}(U^l \cup U^u, \mathscr{T}, \widetilde{\mathcal{FD}})$ make $\mathcal{R}(\widetilde{\mathcal{FC}}_r)(w_t) > 0$. By Theorem 3.1 it is known that $w_t \in CS(\mathscr{V})$ and $w_t \in CS(\mathscr{T})$; hence $CS(\mathscr{V}) \subseteq CS(\mathscr{T})$. Since $\mathscr{T} \subseteq \mathscr{V}$, then $CS(\mathscr{T}) \subseteq CS(\mathscr{V})$, so $CS(\mathscr{T}) = CS(\mathscr{V})$. And because $\mathscr{T}$ is a minimal subset satisfying the condition that for $\forall \widetilde{\mathcal{FC}}_j \in \mathscr{T}$, $CR(\mathscr{T}) \neq CR(\mathscr{T} - \{\widetilde{\mathcal{FC}}_j\})$, then $\mathscr{T}$ is independent. Thus $\mathscr{T}$ is a reduct of $\mathscr{V}$. □

Theorem 3.2 presents reduction rules based on the fuzzy related family. Two new significance functions, grounded in the fuzzy related family, are defined to evaluate the features.

Definition 3.6. Given a FLCIS $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$, where $\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \cdots, \widetilde{\mathcal{FC}}_n\}$ is a fuzzy covering family of $U^l \cup U^u$, $\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$ is the corresponding fuzzy related family. The significance of feature subset $\mathscr{T} \subseteq \mathscr{V}$ is defined as:

$$CD(\mathscr{T}) = \sum_{t=1}^{|U^l \cup U^u|} max\{\mathcal{R}(\widetilde{\mathcal{FC}}_i)(w_t) | \widetilde{\mathcal{FC}}_i \in \mathscr{V}\} \tag{13}$$

For any $\widetilde{\mathcal{FC}}_i \in \mathscr{T}$, the inner significance of $\widetilde{\mathcal{FC}}_i$ with respect to $\mathscr{T}$ is

$$SIG_{in}(\mathscr{T}, \widetilde{\mathcal{FC}}_i) = CD(\mathscr{T}) - CD(\mathscr{T} - \{\widetilde{\mathcal{FC}}_i\}) \tag{14}$$

For any $\widetilde{\mathcal{FC}}_i \in (\mathscr{V} - \mathscr{T})$, the outer significance of $\widetilde{\mathcal{FC}}_i$ with respect to $\mathscr{T}$ is

$$SIG_{out}(\mathscr{T}, \widetilde{\mathcal{FC}}_i) = CD(\mathscr{T} \cup \{\widetilde{\mathcal{FC}}_i\}) - CD(\mathscr{T}) \tag{15}$$

In accordance with 3.6, the inner and outer significance of all features can be calculated using a fuzzy related family without redundant calculations based on the original data or fuzzy label information system. This strategy reduces the time required for reduct computation.

4. Semi-supervised feature selection algorithm based on fuzzy related family

This section introduces a Semi-supervised Feature selection algorithm based on fuzzy Related Family (short for SFRF). The flow chart of SFRF is shown in Fig. 1.

As depicted in Fig. 1, SFRF initially assigns fuzzy labels to unlabeled samples according to 3.1 and computes the fuzzy covering for each conditional feature separately. This process converts the original partially labeled data into a fuzzy label covering information system FLCIS $(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}})$. Subsequently, the fuzzy related family is constructed. Lastly, the significances of feature subsets are determined based on the fuzzy related family, and features are selected following a greedy strategy.

In this paper, the set of fuzzy information granules (fuzzy covering) for each conditional feature a_i is calculated according to the Equation (16) $\widetilde{\mathcal{FC}}_i = \{\widetilde{F}_t | t = 1, 2, \cdots, |U^l|\}$. Then use these fuzzy coverings to form a fuzzy covering family $\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \cdots, \widetilde{\mathcal{FC}}_n\}$.

$$\widetilde{F}_t(w_j) = \begin{cases} \frac{1-|a_i(w_t)-a_i(w_j)|}{\delta}, & |a_i(w_t) - a_i(w_j)| \le \delta \\ 0, & |a_i(w_t) - a_i(w_j)| > \delta \end{cases} \tag{16}$$

where $w_t \in U^l$, $w_j \in U^l \cup U^u$, $\delta \in [0, 0.2]$ is the radius parameters controlling the size of the fuzzy information granules.

For Algorithm 1, the time complexity of adding fuzzy labels is $O(|AT||U^l||U^u|)$. The time complexity of computing the fuzzy information granule and constructing the fuzzy related family is $O(|AT||U^l||U^l \cup U^u|)$. Furthermore, the time complexity of the feature selection phase is $O(k|AT||U^l \cup U^u|)$, where $k < |AT|$ denotes the number of features in the final reduction. Consequently, the overall time complexity for the SFRF algorithm becomes $O(|AT|(|U^l||U^u| + |U^l \cup U^u||U^l| + k|U^l \cup U^u|))$. Since searching all reducts may require exponential time. To improve the computation efficiency, based on a greedy strategy, Algorithm 1 searches a feature subset (an approximate reduct) which may be not a reduct.

Example 4.1. (following Example 3.2) The fuzzy related family obtained in Example 3.2 is

$$\mathcal{FR}(U^l \cup U^u, \mathscr{V}, \widetilde{\mathcal{FD}}) = \{\mathcal{R}(\widetilde{\mathcal{FC}}_1), \mathcal{R}(\widetilde{\mathcal{FC}}_2), \mathcal{R}(\widetilde{\mathcal{FC}}_3), \mathcal{R}(\widetilde{\mathcal{FC}}_4)\}$$

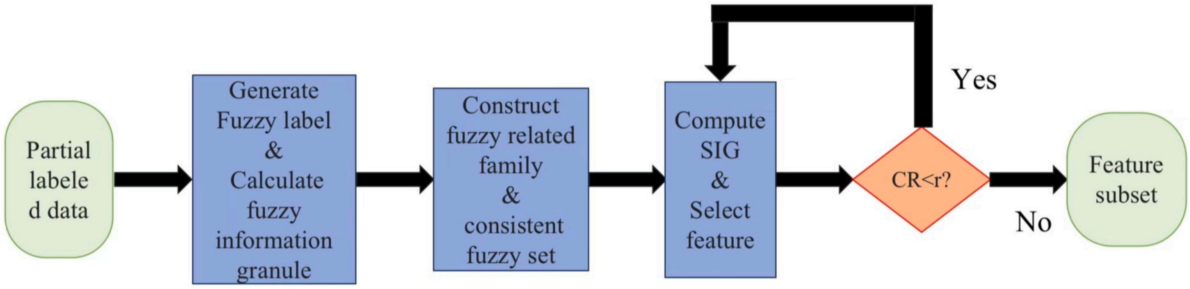


Fig. 1. Flow chart of SFRF algorithm.

Algorithm 1: Semi-supervised feature selection algorithm based on fuzzy related family (SFRF).

Input: A partial label data information system $(U^l \cup U^u, AT, D)$, where $U^l \cup U^u$ is a universe, $AT = \{a_1, a_2, \cdots, a_n\}$ is the conditional feature set, $D = \{d\}$ is the label feature; a radius parameter δ;
Output: An feature subset red;
Generate the fuzzy label $\widetilde{\mathcal{FD}} = \{\widetilde{D}_1, \widetilde{D}_2, \cdots, \widetilde{D}_s\}$ based on Definition 3.1
Let $red = \{\}$;
for $a_i \in AT$ **do**
  Compute the fuzzy covering under the conditional feature a_i according to **Equation (16)** $\widetilde{\mathcal{FC}}_i = \{\widetilde{F}_i | t = 1, 2, \cdots, |U^l|\}$;
  Compute $\mathcal{R}(\widetilde{\mathcal{FC}}_i) = \cup\{\widetilde{F} \in \widetilde{\mathcal{FC}}_i | \exists \widetilde{D}_j \in \widetilde{\mathcal{FD}}$ s.t. $\widetilde{F} \subseteq \widetilde{D}_j\}$;
end for
$\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \cdots, \widetilde{\mathcal{FC}}_n\}$;
Compute the fuzzy positive region coverage $CR(\mathscr{V})$, let $r = CR(\mathscr{V})$;
while $CR(red) < r$ **do**
  for $a_i \in AT - red$ **do**
    Compute $SIG_{out}(red, \widetilde{\mathcal{FC}}_i) = CD(red \cup \{\widetilde{\mathcal{FC}}_i\}) - CD(red)$;
  end for
  Select the $a_*(\widetilde{\mathcal{FC}}_*$ is induced by $a_*)$ with the highest significance $SIG_{out}(red, \widetilde{\mathcal{FC}}_*)$;
  Let $red = red \cup a_*$;
  Compute $CR(red)$;
end while
Back to red.

$$\mathcal{R}(\widetilde{\mathcal{FC}}_1) = \frac{0.8}{w_1} + \frac{0.6}{w_2} + \frac{0.9}{w_3} + \frac{0.8}{w_4} + \frac{0.7}{w_5} + \frac{0}{w_6}, \mathcal{R}(\widetilde{\mathcal{FC}}_2) = \frac{0.3}{w_1} + \frac{0.6}{w_2} + \frac{0.2}{w_3} + \frac{0.5}{w_4} + \frac{0.6}{w_5} + \frac{0}{w_6}$$

$$\mathcal{R}(\widetilde{\mathcal{FC}}_3) = \frac{0.7}{w_1} + \frac{0.6}{w_2} + \frac{0.6}{w_3} + \frac{0}{w_4} + \frac{0.7}{w_5} + \frac{0.6}{w_6}, \mathcal{R}(\widetilde{\mathcal{FC}}_4) = \frac{0.5}{w_1} + \frac{0.5}{w_2} + \frac{0}{w_3} + \frac{0.6}{w_4} + \frac{0.2}{w_5} + \frac{0.5}{w_6}$$

Firstly, we calculate the significance of features $\{\widetilde{\mathcal{FC}}_1\}$, $\{\widetilde{\mathcal{FC}}_2\}$, $\{\widetilde{\mathcal{FC}}_3\}$ and $\{\widetilde{\mathcal{FC}}_4\}$ respectively.

$CD(\{\widetilde{\mathcal{FC}}_1\}) = 3.8$, $CD(\{\widetilde{\mathcal{FC}}_2\}) = 2.2$, $CD(\{\widetilde{\mathcal{FC}}_3\}) = 3.2$, $CD(\{\widetilde{\mathcal{FC}}_4\}) = 2.3$.

Then the feature with the greatest significance $\widetilde{\mathcal{FC}}_1$ is added to feature subset as $\mathscr{T} = \{\widetilde{\mathcal{FC}}_1\}$, calculate $CS(\mathscr{T}) = \{w_1, w_2, w_3, w_4, w_5\}$ and $CR(\mathscr{T}) = 0.83$.

Because the coverage rate of the original conditional feature set $\mathscr{V} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_2, \widetilde{\mathcal{FC}}_3, \widetilde{\mathcal{FC}}_4\}$ is $CR(\mathscr{V}) = 1$. Since $CR(\mathscr{T}) < CR(\mathscr{V})$, we proceed to calculate the significance of the unselected features $\{\widetilde{\mathcal{FC}}_2\}$, $\{\widetilde{\mathcal{FC}}_3\}$ and $\{\widetilde{\mathcal{FC}}_4\}$ respectively.

$SIG_{out}(\mathscr{T}, \{\widetilde{\mathcal{FC}}_2\}) = 0$, $SIG_{out}(\mathscr{T}, \{\widetilde{\mathcal{FC}}_3\}) = 0.6$, $SIG_{out}(\mathscr{T}, \{\widetilde{\mathcal{FC}}_4\}) = 0.5$.

Then the feature with the greatest significance $\widetilde{\mathcal{FC}}_3$ is chosen, $\mathscr{T} = \mathscr{T} \cup \{\widetilde{\mathcal{FC}}_3\} = \{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_3\}$, $CS(\mathscr{T}) = \{w_1, w_2, w_3, w_4, w_5, w_6\}$ and $CR(\mathscr{T}) = 1$. Once $CR(\mathscr{T})$ reach 1, $\{\widetilde{\mathcal{FC}}_1, \widetilde{\mathcal{FC}}_3\}$ is output as the selected feature subset.

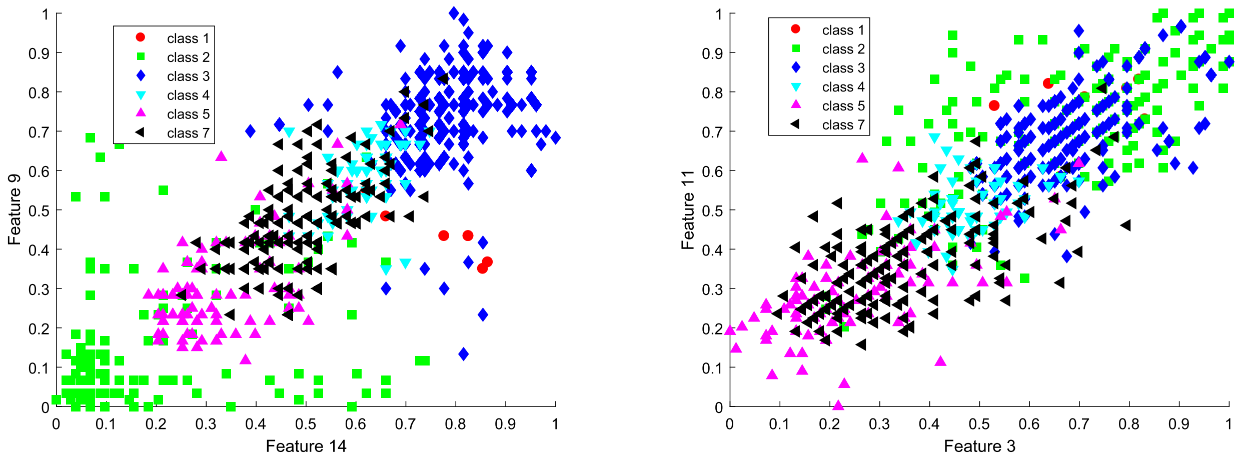
5. Experiment and analysis

In this section, we conducted numerical experiments to verify the effectiveness of SFRF and compared it with three existing semi-supervised feature selection algorithms. These three algorithms were: (1) Neighborhood Granulation based Attribute Reduction for partially labeled decision systems (NGAR) [47]; (2) Semi-supervised Feature selection with fuzzy RElevance and rEdundancy (SEMIFREE) [19]; and (3) Semi-supervised feature selection algorithm combining neighborhood discrimination index and Laplace scoring (Semi-Supervised Neighborhood Discrimination Index, SSNDI) [17].

In terms of parameter setting, both NGAR and SSNDI employ the parameter δ. For these algorithms, we adjusted δ within the range of 0.05 to 0.5, with a step of 0.05. In the case of SFRF, δ was modified between 0.01 to 0.2, with a step of 0.01. Moreover, we adopted the recommended values in [17] for other parameters of SSNDI: $k = 3$, $\lambda = 0.5$, and $\lfloor n/2 \rfloor$, where n signifies the sample size. For every distinct δ value, a feature subset was acquired and the final classification result corresponds to the maximum classification accuracy achieved across the different δ values. In the SEMIFREE, the parameter generating the fuzzy similarity relation, ϵ, was set at 0.2, while the parameter managing the sharpening strength, T, was fixed at 0.5.

Table 3
Data Description.

Index	Dataset	Sample	Feature	Class
1	Seeds	210	7	3
2	Algerian	243	13	2
3	Heart	299	12	2
4	Wholesale	440	7	3
5	Thoracic	470	16	2
6	Ultrasonic	361	43	4
7	Wine	178	13	3
8	Dyrskjot	40	1023	3
9	Lapointe	69	1625	3
10	Risinger	174	1571	10
11	Tomlins	104	2315	5
12	Satimage	6345	36	7



(a) The top two features selected by SFRF (b) Randomly selected two features

Fig. 2. Data distribution of $Satimage$.

The datasets used in the experiments were obtained from the publicly available databases UCI (http://archive.ics.uci.edu/ml) and KEEL (http://sci2s.ugr.es/keel/data). Each dataset was normalized before the experiment, with specific descriptions provided in Table 3.

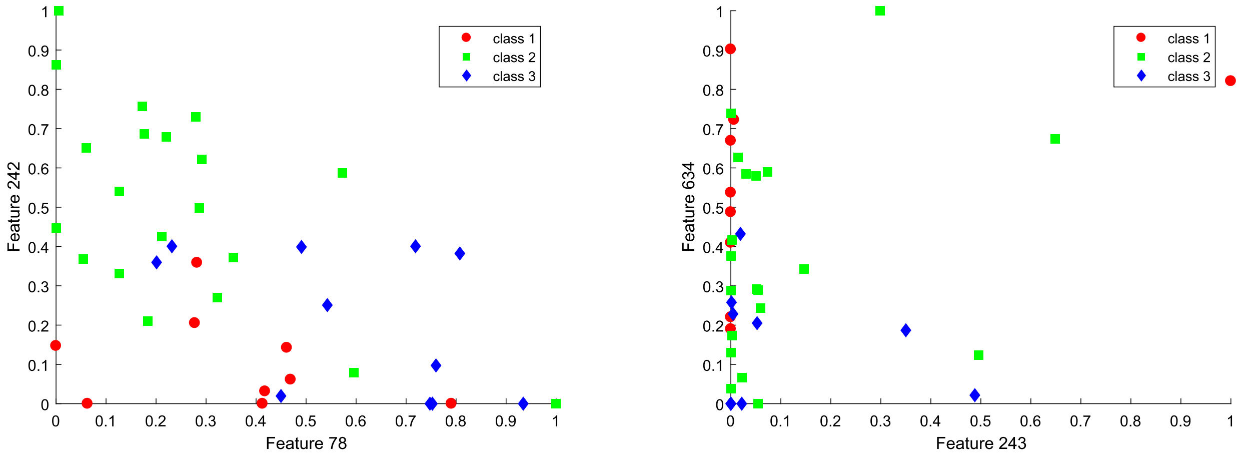
5.1. Classification performance

This subsection primarily demonstrates the classification capability of the feature subset selected by SFRF through numerical experiments.

Firstly, two-dimensional scatter diagrams drew by the first two features of the feature subset are used to illustrate the classification ability of the selected features. Fig. 2, Fig. 3, Fig. 4 and Fig. 5 show the results of the $Satimage$, $Dyrskjot$, $Seeds$, and $Algerian$ datasets (labeling rate 50%, parameter $\delta = 0.05$), and two features are randomly selected for comparison.

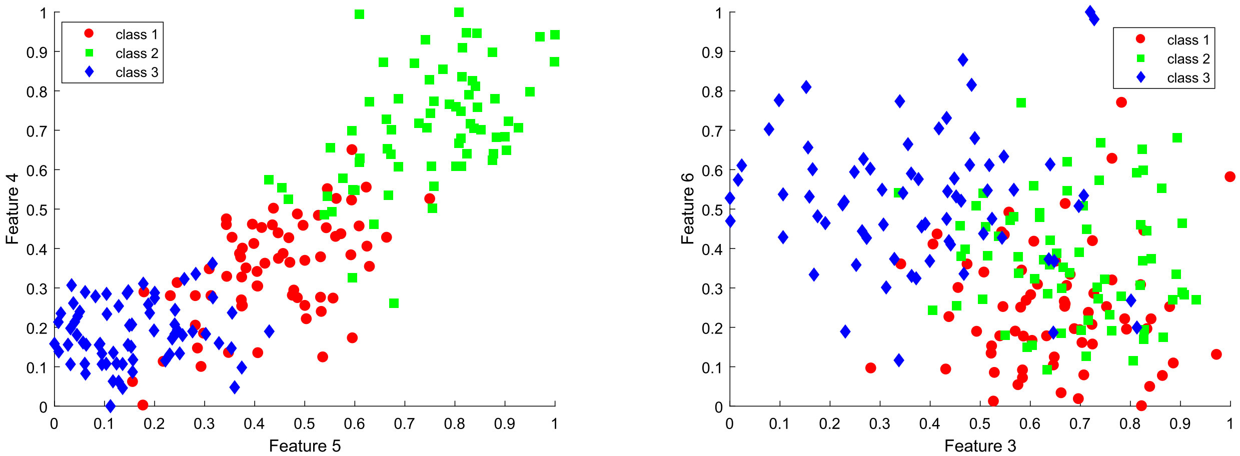
As demonstrated in the figures, the two features selected by SFRF exhibit a stronger ability to distinguish different categories compared to the two randomly selected features. For instance, in the scatter plot of the Satimage dataset (Fig. 2), the ability of the features selected by SFRF to distinguish between different categories is significantly stronger than that of randomly selected features, especially class 2 and class 3, which can be clearly differentiated. In the scatter plot of the Dyrskjot dataset (Fig. 3), although the features selected by SFRF did not distinctly separate the three categories, it is a significant improvement compared to the scenario where the three categories in the randomly selected features are completely mixed together. In the scatter plot of the Seeds dataset (Fig. 4) it can be clearly seen that the primary distribution areas of the three categories are distinctly separated with only minor overlap. In the scatter plot of the Algerian dataset (Fig. 5), the red points and green points representing two categories are almost entirely differentiated.

Next, the classification performances are compared by kNN and CART classifiers. The datasets in Table 3 are all fully labeled. In the experiment, the labeling rate was set at $10\%, 20\%, \ldots, 90\%$ to simulate partially labeled data. For each dataset, a portion of the samples was randomly selected to retain the labels, while the labels of the remaining samples were removed. The four algorithms were then applied to select features for these partially labeled datasets. The results were tested for classification through ten-fold cross-validation, with the classification accuracies displayed as the average of ten classification tests.



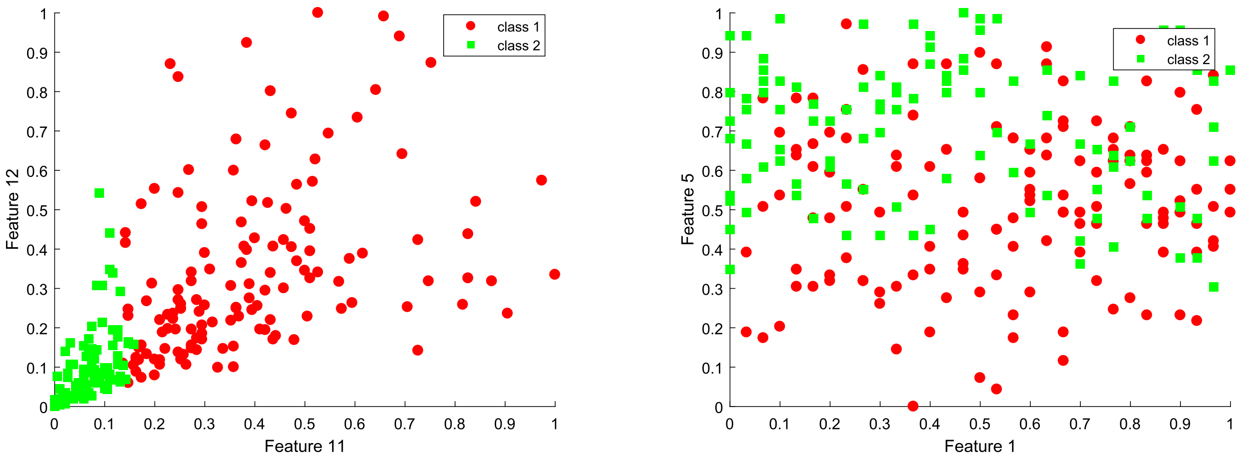
(a) The top two features selected by SFRF (b) Randomly selected two features

Fig. 3. Data distribution of *Dyrskjot*.



(a) The top two features selected by SFRF (b) Randomly selected two features

Fig. 4. Data distribution of *Seeds*.



(a) The top two features selected by SFRF (b) Randomly selected two features

Fig. 5. Data distribution of *Algerian*.

Table 4
Classification accuracy comparison of four algorithms with the labeling rate of 10%, 20% and 30% (bold font indicates the best results among the four algorithms, "NA" represents that the algorithm is unable to complete calculations on this data).

10%	kNN				CART			
Index	SSNDI	SEMIFREE	NGAR	SFRF	SSNDI	SEMIFREE	NGAR	SFRF
1	90.00	**92.86**	**92.86**	90.48	90.00	91.91	**93.81**	91.91
2	95.90	95.87	92.98	**96.73**	**98.77**	**98.77**	**98.77**	96.27
3	72.15	75.60	78.23	**78.56**	71.57	79.38	78.92	**79.58**
4	63.45	71.84	61.17	**71.85**	61.13	71.84	63.22	**71.85**
5	82.53	**83.83**	82.77	82.13	**83.19**	78.94	77.45	77.66
6	**89.50**	82.00	80.06	79.27	85.91	85.86	**87.54**	79.54
7	66.52	**71.52**	71.23	66.84	64.63	63.35	65.20	**65.77**
8	NA	**85.50**	76.00	67.17	NA	**80.50**	63.50	72.67
9	NA	**71.19**	57.68	66.91	NA	**74.70**	67.14	61.43
10	NA	NA	49.57	**71.60**	NA	NA	43.09	**68.45**
11	NA	NA	**63.07**	60.15	NA	NA	50.17	**63.56**
12	NA	NA	NA	**99.05**	NA	NA	NA	**98.23**
20%	kNN				CART			
Index	SSNDI	SEMIFREE	NGAR	SFRF	SSNDI	SEMIFREE	NGAR	SFRF
1	83.81	91.43	92.38	**93.81**	85.24	91.91	**94.76**	90.00
2	96.28	**96.32**	92.16	93.85	98.73	98.75	98.48	**98.78**
3	72.23	**84.65**	80.28	81.31	70.62	80.89	76.23	**80.90**
4	64.11	**71.83**	63.12	62.12	60.94	**71.83**	59.32	59.55
5	81.70	81.70	82.13	**84.26**	82.34	77.87	79.36	**84.26**
6	**87.79**	82.27	80.79	77.53	86.96	86.13	**88.91**	78.66
7	64.92	69.32	70.41	**72.60**	**69.28**	62.49	64.69	66.03
8	NA	**87.17**	65.50	62.83	NA	**87.50**	62.50	83.50
9	NA	**74.11**	71.43	65.24	NA	**79.40**	63.63	66.91
10	NA	NA	43.40	**76.39**	NA	NA	35.69	**65.84**
11	NA	NA	59.70	**73.00**	NA	NA	49.10	**56.91**
12	NA	NA	NA	**99.01**	NA	NA	NA	**98.24**
30%	kNN				CART			
Index	SSNDI	SEMIFREE	NGAR	SFRF	SSNDI	SEMIFREE	NGAR	SFRF
1	**94.76**	93.81	93.81	92.86	92.86	92.86	**94.29**	**94.29**
2	**97.13**	96.21	91.70	91.78	**99.18**	98.78	98.75	98.80
3	70.92	**85.61**	82.53	83.22	70.15	**80.27**	79.24	72.73
4	63.00	**71.82**	63.44	61.84	60.48	**71.82**	61.05	62.72
5	80.85	81.49	**82.13**	81.70	**82.98**	78.30	78.09	77.02
6	87.54	87.76	81.19	**88.91**	84.24	87.85	85.05	**88.35**
7	68.76	70.13	70.10	**70.75**	66.91	66.58	67.10	**67.16**
8	NA	85.83	68.17	**95.50**	NA	**89.17**	66.50	76.00
9	NA	73.93	71.49	**75.48**	NA	**76.01**	70.48	65.71
10	NA	NA	48.66	**81.12**	NA	NA	41.20	**74.17**
11	NA	NA	62.51	**78.02**	NA	NA	59.47	**57.68**
12	NA	NA	NA	**98.99**	NA	NA	NA	**98.26**

Table 4, Table 5 and Table 6 present the classification accuracy of SFRF and the three comparative algorithms NGAR, SSNDI, and SEMIFREE evaluated by the kNN ($k = 3$) and CART classifiers at various label rates. The performance of SFRF was analyzed at different label rates, and the results are summarized as follows: At the 10% label rate, SFRF exhibited the highest accuracy among the four algorithms on 5 and 6 datasets when tested with kNN and CART classifiers, respectively (out of a total of 12 datasets). At 20%, SFRF led in accuracy on 6 datasets for both classifiers. At 30%, the highest accuracy was achieved on 8 and 6 datasets with kNN and CART, respectively. At 40%, SFRF outperformed on 4 and 7 datasets with kNN and CART, respectively. At 50%, the algorithm led in accuracy on 7 and 6 datasets with kNN and CART, respectively. At 60%, SFRF maintained the highest accuracy on 7 and 6 datasets for kNN and CART, respectively. At 70%, the algorithm achieved leading accuracy on 9 and 8 datasets with kNN and CART, respectively. At 80%, SFRF was the top performer on 8 datasets for both classifiers. Finally, at a 90% label rate, SFRF achieved the highest accuracy on 8 and 9 datasets with kNN and CART, respectively.

Taking into account the results from Table 4, Table 5 and Table 6, regardless of the kNN classifier or the CART classifier, SFRF can achieve comparable, and even better, classification results compared to the contrast algorithms.

5.2. Computational efficiency

In this section, the computational efficiency of SFRF is compared with three existing algorithms. The total execution time of the four algorithms for 9 different labeling rates is presented in Table 7, while the execution time of a single round for different labeling rates is illustrated in Fig. 6. The maximum execution time for all algorithms was set as 12 hours.

As can be seen from Table 7, SFRF achieved the shortest execution time on eight of the test datasets. In the remaining four instances where SFRF didn't secure the fastest time, they are all small-scale datasets, SFRF was second only to SEMIFREE. The 8th through the 11th datasets, which feature medium-to-high dimensionality, displayed a significant surge in the execution time for

Table 5
Classification accuracy comparison of four algorithms with the labeling rate of 40%, 50% and 60% (bold font indicates the best results among the four algorithms, “NA” represents that the algorithm is unable to complete calculations on this data).

40%	*k*NN				CART			
Index	SSNDI	SEMIFREE	NGAR	SFRF	SSNDI	SEMIFREE	NGAR	SFRF
1	**94.76**	94.29	91.91	92.86	93.33	91.91	**94.29**	93.33
2	95.08	**95.87**	90.98	92.65	97.95	**98.78**	98.38	**98.78**
3	72.29	82.22	**80.61**	86.00	69.56	80.54	77.21	**80.98**
4	62.51	**71.83**	62.26	62.97	61.83	**71.83**	60.22	61.38
5	81.49	**83.19**	81.49	79.36	81.70	80.00	78.51	**83.62**
6	**89.74**	85.38	81.02	86.18	85.66	88.13	85.04	**90.04**
7	67.12	70.09	**73.15**	69.87	**67.91**	66.26	63.86	63.78
8	NA	78.17	73.17	**78.33**	NA	**93.00**	79.67	88.50
9	NA	**78.69**	67.14	61.13	NA	**75.36**	63.81	63.99
10	NA	NA	55.63	**79.46**	NA	NA	50.61	**79.49**
11	NA	NA	73.89	**76.38**	NA	NA	54.78	**62.96**
12	NA	NA	NA	**99.10**	NA	NA	NA	**98.24**

50%	*k*NN				CART			
Index	SSNDI	SEMIFREE	NGAR	SFRF	SSNDI	SEMIFREE	NGAR	SFRF
1	**94.29**	92.86	92.38	92.38	**94.29**	92.86	**94.29**	93.33
2	95.01	95.83	92.16	**96.33**	98.75	99.17	98.77	**99.18**
3	74.95	82.96	78.60	**84.30**	69.23	78.98	75.21	**79.98**
4	62.98	**71.83**	63.24	62.30	61.58	**71.83**	59.32	62.33
5	81.49	**82.34**	81.49	**82.34**	**82.55**	79.57	79.15	77.87
6	**86.48**	84.79	80.42	76.93	86.19	86.72	85.86	**87.82**
7	66.59	67.43	70.44	**71.21**	**67.36**	66.40	66.57	62.98
8	NA	**80.67**	65.67	75.17	NA	81.00	62.67	**84.00**
9	NA	**79.88**	63.81	70.83	NA	**84.17**	75.48	69.23
10	NA	NA	51.79	**82.70**	NA	NA	44.17	**72.47**
11	NA	NA	63.27	**79.36**	NA	NA	**57.55**	56.75
12	NA	NA	NA	**99.07**	NA	NA	NA	**98.42**

60%	*k*NN				CART			
Index	SSNDI	SEMIFREE	NGAR	SFRF	SSNDI	SEMIFREE	NGAR	SFRF
1	**94.29**	**94.29**	92.86	93.81	**93.81**	91.43	93.33	**93.81**
2	**95.90**	**95.90**	93.03	93.38	**98.77**	**98.77**	**98.77**	98.75
3	71.63	82.97	**82.96**	**82.96**	69.92	**82.34**	75.23	79.29
4	63.26	**71.83**	61.86	63.85	66.17	**71.83**	61.21	60.86
5	81.06	81.28	**82.34**	**82.34**	**82.55**	81.28	80.00	79.15
6	86.18	87.78	80.63	**89.81**	84.21	85.93	85.62	**88.90**
7	68.81	69.89	**71.48**	67.63	66.01	65.83	66.05	**70.14**
8	NA	**88.00**	71.00	86.67	NA	**88.50**	69.00	79.17
9	NA	75.48	66.25	**82.74**	NA	**81.19**	65.95	70.77
10	NA	NA	49.90	**82.96**	NA	NA	51.24	**82.33**
11	NA	NA	59.75	**77.74**	NA	NA	56.87	**61.87**
12	NA	NA	NA	**99.07**	NA	NA	NA	**98.28**

SEMIFREE, which initially showed the shortest runtime on the first quartet of datasets. For datasets 10 through 12, SEMIFREE even overshooted the maximum execution time. For datasets spanning from 8 to 12, SSNDI persistently breached the maximum execution time, and NGAR surpassed this threshold on the 12th dataset. These results collectively testify to efficiency of SFRF as a feature selection algorithm.

Fig. 6 presents the execution time according to different label rates, with each subplot comparing the performance of SFRF with three other algorithms. It is distinctly noticeable that SFRF maintains its high efficiency consistently. Furthermore, in subplots h, i, j, and k, representing medium-to-high dimensional datasets, other algorithms show significant volatility, whereas SFRF continues to exhibit robust performance.

In an overarching view, SFRF holds the shortest cumulative time among all four algorithms. Despite performance of SFRF being slightly superseded by the SEMIFREE algorithm on small scale datasets, it surpasses all other comparison algorithms by a substantial margin when applied to larger scale datasets. This underlines that the newly proposed algorithm considerably escalates the efficiency of feature selection, thereby signifying its suitability for the processing of large-scale datasets.

6. Conclusion and future work

In this paper, we propose a semi-supervised feature selection method called fuzzy related family based on fuzzy rough sets, which generates fuzzy labels for unlabeled samples using fuzzy relationships to address the problem of semi-supervised feature selection for partially labeled data. Based on the fuzzy related family, an efficient semi-supervised feature selection algorithm, SFRF, is designed. Experimental results demonstrate the effectiveness of the proposed method on data with various labeling rates. Compared with three existing semi-supervised feature selection algorithms, while preserving the consistent level of classification accuracy, SFRF

Table 6
Classification accuracy comparison of four algorithms with the labeling rate of 70%, 80% and 90% (bold font indicates the best results among the four algorithms, "NA" represents that the algorithm is unable to complete calculations on this data).

70%	*k*NN				CART			
Index	SSNDI	SEMIFREE	NGAR	SFRF	SSNDI	SEMIFREE	NGAR	SFRF
1	**94.76**	92.86	93.81	93.33	93.33	90.95	**94.76**	94.29
2	95.86	91.77	93.01	**95.92**	**98.78**	**98.78**	**98.78**	**98.78**
3	73.28	69.51	83.92	**84.56**	71.21	71.00	78.55	**79.89**
4	64.12	**71.84**	61.40	62.76	59.30	**71.84**	61.13	59.51
5	82.34	81.06	79.79	**83.62**	82.13	81.28	81.91	80.00
6	**89.76**	87.29	81.18	89.04	85.03	88.28	85.41	**88.67**
7	67.13	71.49	71.54	**72.38**	67.19	65.72	62.52	63.78
8	NA	86.00	85.50	**92.50**	NA	85.50	68.50	**87.50**
9	NA	66.67	68.99	**70.71**	NA	69.23	64.76	**86.85**
10	NA	NA	48.09	**81.46**	NA	NA	40.52	**72.56**
11	NA	NA	66.93	**75.66**	NA	NA	48.28	**66.47**
12	NA	NA	NA	**99.08**	NA	NA	NA	**98.24**
80%	*k*NN				CART			
Index	SSNDI	SEMIFREE	NGAR	SFRF	SSNDI	SEMIFREE	NGAR	SFRF
1	**93.81**	93.33	93.33	92.38	93.33	93.33	92.86	**93.81**
2	95.07	90.91	94.27	**95.85**	97.48	**98.75**	96.73	98.73
3	71.62	68.36	81.25	**84.91**	72.30	69.64	78.56	**79.95**
4	61.83	**71.83**	61.21	64.57	62.00	**71.83**	61.61	63.46
5	81.49	80.64	81.91	**82.55**	**81.70**	81.06	80.21	78.30
6	**88.12**	86.20	80.60	80.59	86.50	86.48	86.66	**87.50**
7	66.34	69.86	**70.72**	71.50	67.13	63.86	63.28	**68.22**
8	NA	**85.00**	77.67	**85.00**	NA	77.50	**83.00**	77.50
9	NA	67.08	66.91	**82.50**	NA	70.83	65.89	**85.83**
10	NA	NA	54.53	**84.46**	NA	NA	50.33	**73.50**
11	NA	NA	60.56	**80.73**	NA	NA	53.83	**60.18**
12	NA	NA	NA	**98.99**	NA	NA	NA	**98.37**
90%	*k*NN				CART			
Index	SSNDI	SEMIFREE	NGAR	SFRF	SSNDI	SEMIFREE	NGAR	SFRF
1	93.33	92.86	92.38	**93.81**	93.81	92.38	**94.29**	**94.29**
2	**96.26**	90.15	91.78	95.45	**98.77**	**98.77**	**98.77**	**98.77**
3	72.30	68.24	81.24	**84.26**	69.93	71.00	76.56	**81.19**
4	62.92	**71.83**	62.03	63.23	60.72	**71.83**	60.51	61.81
5	80.85	81.06	82.77	**82.98**	**82.77**	81.06	79.79	80.21
6	**88.38**	80.19	82.22	83.42	87.26	**87.82**	85.29	85.59
7	64.72	70.72	**72.89**	67.64	66.55	63.85	64.93	**67.65**
8	NA	85.00	87.17	**88.00**	NA	85.50	85.00	**92.50**
9	NA	65.30	78.27	**86.07**	NA	71.31	62.74	**83.39**
10	NA	NA	56.48	**90.82**	NA	NA	46.89	**80.65**
11	NA	NA	70.60	**75.29**	NA	NA	55.92	**63.87**
12	NA	NA	NA	**99.05**	NA	NA	NA	**98.09**

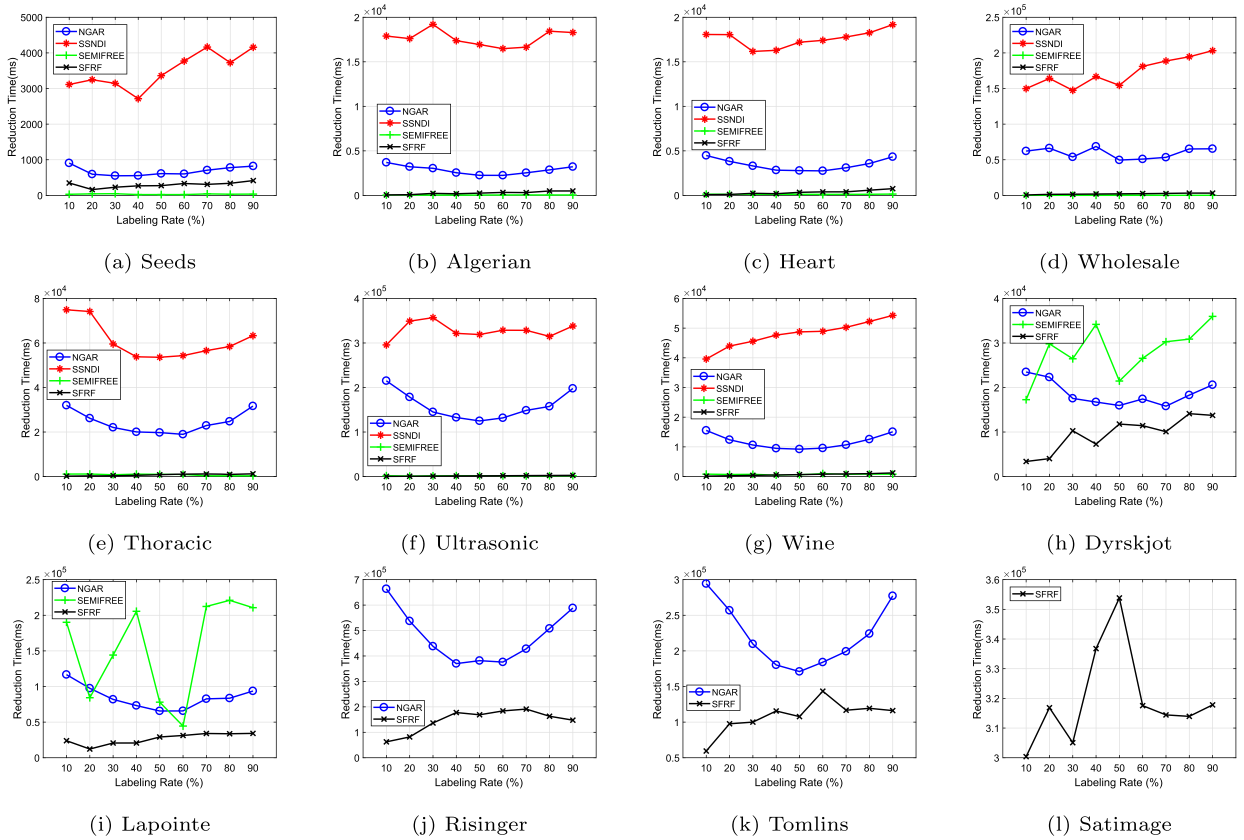
Table 7
The total execution time of the four algorithms for 9 different labeling rates (in milliseconds).

Index	SSNDI	SEMIFREE	NGAR	SFRF
1	31399	298	6125	2676
2	158857	359	25601	2418
3	158403	1144	30974	3067
4	1550135	800	535318	18997
5	548522	7059	218289	6668
6	2952019	14850	1431698	13601
7	431253	6687	105142	5879
8	NA	252816	168134	86125
9	NA	1390617	760518	239791
10	NA	NA	4290993	1313412
11	NA	NA	1996654	976264
12	NA	NA	NA	2876627

considerably elevates the computational velocity, a noteworthy advancement. As we extend our research in subsequent investigations, we are committed to endeavoring the translation of this algorithm into practical, real-world application scenarios.

CRediT authorship contribution statement

Zhijun Guo (1st author): Paper writing, theoretical derivation and algorithm designing;



(a) Seeds (b) Algerian (c) Heart (d) Wholesale

(e) Thoracic (f) Ultrasonic (g) Wine (h) Dyrskjot

(i) Lapointe (j) Risinger (k) Tomlins (l) Satimage

Fig. 6. Execution Time of reduction with different labeling rate (in milliseconds).

Tian Yang (the corresponding author, 2nd author): Theoretical framework constructing, paper writing, paper submission;
Yuanjiang Li (3rd author): Paper writing, algorithm designing, experimental analysis;
Yang Shen (4th author): Paper writing, experimental analysis, data collecting;
Yanfang Deng (5th author): Paper writing, algorithm modification;
Yuhua Qian (6th author): Theoretical framework constructing, algorithm designing.

Declaration of competing interest

There is not any conflict of interest.

Data availability

Data will be made available on request.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (No. 11201490 and No. 61976089), the Natural Science Foundation of Hunan Province under Grant (No. 2021JJ20037), the Training Program for Excellent Young Innovators of Changsha (No. kq1905031), the National Key Research and Development Program of China (No. 2021ZD0112400), the Key Program of the National Natural Science Foundation of China (No. 62136005).

References

[1] X. Wu, H. Chen, T. Li, J. Wan, Semi-supervised feature selection with minimal redundancy based on local adaptive, Appl. Intell. 51 (2021) 8542–8563.
[2] J. Lai, H. Chen, T. Li, X. Yang, Adaptive graph learning for semi-supervised feature selection with redundancy minimization, Inf. Sci. 609 (2022) 465–488.
[3] B. Jiang, X. Wu, X. Zhou, Y. Liu, A.G. Cohn, W. Sheng, H. Chen, Semi-supervised multiview feature selection with adaptive graph learning, IEEE Trans. Neural Netw. Learn. Syst. (2022), https://doi.org/10.1109/TNNLS.2022.3194957.
[4] S. Lv, S. Shi, H. Wang, F. Li, Semi-supervised multi-label feature selection with adaptive structure learning and manifold learning, Knowl.-Based Syst. 214 (2021) 106757.

[5] S. An, M. Zhang, C. Wang, W. Ding, Robust fuzzy rough approximations with knn granules for semi-supervised feature selection, Fuzzy Sets Syst. 461 (2023) 108476.
[6] W. Shu, Z. Yan, J. Yu, W. Qian, Information gain-based semi-supervised feature selection for hybrid data, Appl. Intell. 53 (6) (2023) 7310–7325.
[7] A. Campagner, D. Ciucci, E. Hüllermeier, Rough set-based feature selection for weakly labeled data, Int. J. Approx. Reason. 136 (2021) 150–167.
[8] A. Campagner, D. Ciucci, T. Denœux, Belief functions and rough sets: survey and new insights, Int. J. Approx. Reason. 143 (2022) 192–215.
[9] A. Campagner, D. Ciucci, Rough-set based genetic algorithms for weakly supervised feature selection, in: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Springer, 2022, pp. 761–773.
[10] K. Liu, X. Yang, H. Yu, J. Mi, P. Wang, X. Chen, Rough set based semi-supervised feature selection via ensemble selector, Knowl.-Based Syst. 165 (2019) 282–296.
[11] Z. Xu, I. King, M.R.-T. Lyu, R. Jin, Discriminative semi-supervised feature selection via manifold regularization, IEEE Trans. Neural Netw. 21 (7) (2010) 1033–1047.
[12] Z. Li, J. Tang, Semi-supervised local feature selection for data classification, Sci. China Inf. Sci. 64 (9) (2021) 192108.
[13] Y. Qian, X. Liang, Q. Wang, J. Liang, B. Liu, A. Skowron, Y. Yao, J. Ma, C. Dang, Local rough set: a solution to rough data analysis in big data, Int. J. Approx. Reason. 97 (2018) 38–63.
[14] Q. Wang, Y. Qian, X. Liang, Q. Guo, J. Liang, Local neighborhood rough set, Knowl.-Based Syst. 153 (2018) 53–64.
[15] T. Yang, Y. Deng, B. Yu, Y. Qian, J. Dai, Local feature selection for large-scale data sets limited labels, IEEE Trans. Knowl. Data Eng. 35 (2022) 7152–7163.
[16] J. Dai, Q. Hu, J. Zhang, H. Hu, N. Zheng, Attribute selection for partially labeled categorical data by rough set approach, IEEE Trans. Cybern. 47 (9) (2016) 2460–2471.
[17] Q. Pang, L. Zhang, Semi-supervised neighborhood discrimination index for feature selection, Knowl.-Based Syst. 204 (2020) 106224.
[18] W. Shu, J. Yu, Z. Yan, W. Qian, Semi-supervised feature selection for partially labeled mixed-type data based on multi-criteria measure approach, Int. J. Approx. Reason. 153 (2023) 258–279.
[19] K. Liu, T. Li, X. Yang, H. Chen, J. Wang, Z. Deng, Semifree: semi-supervised feature selection with fuzzy relevance and redundancy, IEEE Trans. Fuzzy Syst. (2023), https://doi.org/10.1109/TFUZZ.2023.3255893.
[20] W. Qian, Y. Li, Q. Ye, W. Ding, W. Shu, Disambiguation-based partial label feature selection via feature dependency and label consistency, Inf. Fusion 94 (2023) 152–168.
[21] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[22] K. Liu, X. Yang, H. Yu, J. Mi, P. Wang, X. Chen, Rough set based semi-supervised feature selection via ensemble selector, Knowl.-Based Syst. 165 (2019) 282–296.
[23] J. Zhao, J. ming Liang, Z. ning Dong, D. yu Tang, Z. Liu, Nec: a nested equivalence class-based dependency calculation approach for fast feature selection using rough set theory, Inf. Sci. 536 (2020) 431–453.
[24] H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (8) (2005) 1226–1238.
[25] P. Sowkuntla, P.S. Prasad, Mapreduce based parallel fuzzy-rough attribute reduction using discernibility matrix, Appl. Intell. 52 (1) (2022) 154–173.
[26] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (3) (1965) 338–353.
[27] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, Int. J. Gen. Syst. 17 (2–3) (1990) 191–209.
[28] W. Li, H. Zhou, W. Xu, X.-Z. Wang, W. Pedrycz, Interval dominance-based feature selection for interval-valued ordered data, IEEE Trans. Neural Netw. Learn. Syst. (2022), https://doi.org/10.1109/TNNLS.2022.3184120.
[29] W. Xu, K. Yuan, W. Li, W. Ding, An emerging fuzzy feature selection method using composite entropy-based uncertainty measure and data distribution, IEEE Trans. Emerg. Top. Comput. Intell. (2022), https://doi.org/10.1109/TETCI.2022.3171784.
[30] X. Yang, H. Chen, T. Li, C. Luo, A noise-aware fuzzy rough set approach for feature selection, Knowl.-Based Syst. 250 (2022) 109092.
[31] R.K. Huda, H. Banka, Efficient feature selection methods using pso with fuzzy rough set as fitness function, Soft Comput. (2022) 1–21.
[32] J. Dai, H. Hu, W.-Z. Wu, Y. Qian, D. Huang, Maximal-discernibility-pair-based approach to attribute reduction in fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26 (4) (2017) 2174–2187.
[33] L. Sun, L. Wang, W. Ding, Y. Qian, J. Xu, Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets, IEEE Trans. Fuzzy Syst. 29 (1) (2021) 19–33.
[34] Z. Huang, J. Li, Y. Qian, Noise-tolerant fuzzy-β-covering-based multigranulation rough sets and feature subset selection, IEEE Trans. Fuzzy Syst. 30 (7) (2022) 2721–2735.
[35] Q. Hu, L. Zhang, Y. Zhou, W. Pedrycz, Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets, IEEE Trans. Fuzzy Syst. 26 (1) (2017) 226–238.
[36] M. Ma, T. Deng, N. Wang, Y. Chen, Semi-supervised rough fuzzy Laplacian eigenmaps for dimensionality reduction, Int. J. Mach. Learn. Cybern. 10 (2019) 397–411.
[37] J. Xing, C. Gao, J. Zhou, Weighted fuzzy rough sets-based tri-training and its application to medical diagnosis, Appl. Soft Comput. 124 (2022) 109025.
[38] T. Yang, Q. Li, B. Zhou, Related family: a new method for attribute reduction of covering information systems, Inf. Sci. 228 (2013) 175–191.
[39] T. Yang, X. Zhong, G. Lang, Y. Qian, J. Dai, Granular matrix: a new approach for granular structure reduction and redundancy evaluation, IEEE Trans. Fuzzy Syst. 28 (12) (2020) 3133–3144.
[40] T. Yang, J. Liang, Y. Pang, P. Xie, Y. Qian, R. Wang, An efficient feature selection algorithm based on the description vector and hypergraph, Inf. Sci. 629 (2023) 746–759.
[41] G. Lang, M. Cai, H. Fujita, Q. Xiao, Related families-based attribute reduction of dynamic covering decision information systems, Knowl.-Based Syst. 162 (2018) 161–173.
[42] G. Lang, Q. Li, M. Cai, H. Fujita, H. Zhang, Related families-based methods for updating reducts under dynamic object sets, Knowl. Inf. Syst. 60 (2019) 1081–1104.
[43] T. Feng, S.-P. Zhang, J.-S. Mi, The reduction and fusion of fuzzy covering systems based on the evidence theory, Int. J. Approx. Reason. 53 (1) (2012) 87–103.
[44] L. Ma, Two fuzzy covering rough set models and their generalizations over fuzzy lattices, Fuzzy Sets Syst. 294 (2016) 1–17.
[45] B. Yang, B.Q. Hu, A fuzzy covering-based rough set model and its generalization over fuzzy lattice, Inf. Sci. 367 (2016) 463–486.
[46] T. Yang, X. Zhong, G. Lang, Y. Qian, J. Dai, Granular matrix: a new approach for granular structure reduction and redundancy evaluation, IEEE Trans. Fuzzy Syst. 28 (12) (2020) 3133–3144.
[47] B. Li, J. Xiao, X. Wang, Feature selection for partially labeled data based on neighborhood granulation measures, IEEE Access 7 (2019) 37238–37250.