Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

An Improved Bi-granulation Feature Selection Method for Interval-valued Multi-label Data Based on Fuzzy Rough Sets

Shuwen Zheng
Minnan Normal University
Guoping Lin
guoplin@163.com

Minnan Normal University
Weihua Xu
Minnan Normal University
Yujie Qin
Minnan Normal University

Research Article

Keywords:

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11722336/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

An Improved Bi-granulation Feature Selection Method for Interval-valued Multi-label Data Based on Fuzzy Rough Sets

ZHENG Shuwen[1], LIN Guoping[1,2,3,*], XU Weihua[1,4], QIN Yujie[1]

[1] School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, Fujian, China
[2] Fujian Key Laboratory of Granular Computing and Its Applications, Zhangzhou 363000, Fujian, China
[3] Institute of Meteorological Big Data of Digital Fujian, Minnan Normal University, Zhangzhou 363000, Fujian, China
[4] College of Artificial Intelligence, Southwest University, Chongqing 400715, China

Abstract Interval-valued multi-label data are characterized by uncertainty in conditional attribute values and diversity in decision label semantics, which poses substantial challenges to feature selection. This study develops an improved bi-granulation feature selection approach based on fuzzy rough sets. The proposed method characterizes relationships among samples from both the conditional space and the decision space. In the conditional space, an improved conditional granulation relation is constructed. In the decision space, a label-weighting mechanism is introduced to enhance the supervisory effect of labels. On this foundation, a unified fuzzy rough set framework is constructed, within which the fuzzy positive region, dependency degree, and attribute significance are formulated for feature selection. Empirical comparisons indicate that the method captures the relationship between interval-valued conditional features and multi-label decision information well, and it delivers competitive classification results while remaining stable across several evaluation metrics.

Keywords: granular computing, interval-valued multi-label data, fuzzy rough sets, bi-granulation, feature selection

1 Introduction

In machine learning and data mining tasks, high-dimensional data often contain many redundant features and noisy attributes. How to select a subset of features that is both representative and discriminative from the original feature space is a core issue for improving model performance, reducing computational cost, and enhancing the interpretability of results. Granular computing mainly organizes, abstracts, and represents data at different levels of granularity, and plays an important role in uncertainty modeling, knowledge discovery, and complex problem analysis. Within the framework of granular computing, feature selection involves the screening and optimization of information granules. Therefore, studying feature selection within the framework of granular computing has clear theoretical significance and practical application value.

Among granular computing-related methods, rough set theory serves as a representative approach. It can perform attribute reduction directly without requiring prior information, and has therefore been widely used in feature selection research[1, 2]. A limitation of classical rough sets is that they are built on strict

Corresponding author: LIN Guoping, E-mail: guoplin@163.com

equivalence relations, making them less suitable for continuous, fuzzy, and uncertain data[3]. Dubois and Prade[4] combined fuzzy set concepts with rough sets and thereby introduced the fuzzy rough set model. Later, Wang et al.[5] characterized fuzzy dependency through fuzzified decisions, while Hu et al.[6] extended this line of research by establishing neighborhood rough set theory. On this basis, Wang et al.[7], Xia et al.[8], Peng et al.[9], and Wang et al.[10] further extended rough set models from the perspectives of fuzzy neighborhoods, granular-ball neighborhoods, weighted neighborhoods, and weighted fuzzy approximation operators, respectively, enabling these models to better adapt to feature selection tasks in complex data environments.

With the increasing richness of semantic structures in data objects, single-label learning is no longer sufficient in many scenarios to fully characterize the decision information of samples. In multi-label learning, each sample may correspond to multiple semantic labels. This property makes the paradigm applicable to tasks such as text classification, image annotation, and bioinformatics analysis. For multi-label feature selection, Dai et al.[11] studied the problem from the perspectives of label gain and label collaboration. Deng et al.[12] established a weighted label-based fuzzy relevance and redundancy model. Zou et al.[13] proposed a gradient feature selection method that considers three-way interaction relationships. Mohanraju et al.[14] introduced a multi-scale feature selection approach for multi-label data by exploiting fuzzy mutual information. In addition, Zhang et al.[15] combined label enhancement techniques with mutual information measures to mine latent semantic label information. Yu et al.[16] proposed a dual-perspective class-imbalance-aware feature selection method, while Dai et al.[17] investigated the learning of common and specific labels in multi-view multi-label data using fuzzy mutual information. Overall, these methods have achieved promising results on deterministic multi-label data. However, in interval-valued multi-label scenarios where both attribute uncertainty and label semantic multiplicity coexist, the applicability of existing methods remains insufficient.

In practical applications, sample attributes are not always given as precise values, but are often represented in the form of intervals to describe measurement errors, ranges of variation, and cognitive uncertainty. As interval-valued data are common in uncertain environments, developing effective feature selection techniques for such data has drawn considerable attention in uncertain data analysis. In this context, Qian et al.[18] established a framework for interval-valued ordered information systems. Li et al.[19] presented a feature selection method based on interval dominance relations. Ma et al.[20] focused on semi-supervised feature selection in partially labeled interval-valued information systems. However, when existing methods are directly applied to interval-valued multi-label scenarios, two prominent problems still arise. First, in the conditional space, many methods focus more on the differences in interval center positions while paying insufficient attention to the internal dispersion characteristics of intervals, which affects the fine-grained description of inter-sample relationships. Second, in the decision space, labels are usually assigned equal weights, and the discriminative role of low-frequency labels is not fully reflected, thereby weakening the ability of the decision granulation relation to capture label difference information.

In response to these challenges, a bi-granulation fuzzy rough set model is constructed for interval-valued multi-label data, which simultaneously exploits information from

the conditional space and the decision space to characterize inter-sample relationships. In the conditional space, a composite interval distance integrating the difference in interval midpoint positions and the difference in internal dispersion is constructed, based on which an improved conditional granulation relation is established. In the decision space, a weighted decision granulation relation is constructed by introducing normalized inverse label-frequency weights. Based on these two types of granulation relations, a unified fuzzy rough set model is further established. Based on this model, the fuzzy positive region, dependency degree, and attribute significance are formulated, from which a feature selection algorithm is further derived.

The main contributions of this paper are summarized as follows:

1. In the conditional space, a composite interval distance integrating interval midpoint differences and internal dispersion differences is constructed, and a conditional granulation relation is established.

2. In the decision space, a normalized inverse label-frequency weighting mechanism is introduced to establish a decision granulation relation.

3. By integrating granulation information from the conditional space and the decision space, an interval-valued multi-label fuzzy rough set model is constructed, based on which a feature selection algorithm is designed.

2 Preliminaries

This section reviews the preliminaries needed for the proposed model, including interval-valued multi-label decision systems and fuzzy similarity relations. We first introduce the interval-valued multi-label decision system and the corresponding interval-valued representation. Then, fuzzy similarity relations and their related properties are described, so as to provide the basis for defining granulation relations in both the conditional space and the decision space.

2.1 Interval-valued multi-label decision systems

Definition 2.1 [21] An interval-valued multi-label decision system is represented as a triplet $IV\text{-}MLDS = \langle U, AT, L\rangle$, where $U = \{x_1, x_2, \ldots, x_n\}$ is a nonempty finite sample universe, $AT = \{a_1, a_2, \ldots, a_m\}$ is the set of interval-valued conditional attributes, and $L = \{l_1, l_2, \ldots, l_q\}$ is the set of multi-label decision labels.

Different from a classical single-label decision system, an interval-valued multi-label decision system uses interval values to characterize uncertainty in sample attributes, while the decision part allows the same sample to be associated with multiple labels. Therefore, $IV\text{-}MLDS$ can simultaneously describe the interval-valued attribute characteristics and the multi-label decision characteristics of data objects.

For any sample $x_i \in U$ and attribute $a_k \in AT$, let the value of x_i on attribute a_k be denoted by $a_k(x_i) = [a_k^-(x_i), a_k^+(x_i)]$, where $a_k^-(x_i)$ and $a_k^+(x_i)$ are the lower and upper bounds of the interval, respectively, satisfying $a_k^-(x_i) \le a_k^+(x_i)$. When $a_k^-(x_i) = a_k^+(x_i)$, the interval value degenerates into a precise value.

To characterize the basic properties of interval values, the interval length and the interval midpoint are defined as follows:

$$\mathrm{len}(a_k(x_i)) = a_k^+(x_i) - a_k^-(x_i),$$
$$\mathrm{mid}(a_k(x_i)) = \frac{a_k^-(x_i) + a_k^+(x_i)}{2}.$$

Here, $\mathrm{len}(a_k(x_i))$ represents the dispersion de-

gree of the interval, while $\mathrm{mid}(a_k(x_i))$ indicates the position of the interval on the real axis.

For any sample $x_i \in U$, its label information is represented by a q-dimensional binary vector $Y_i = (y_{i1}, y_{i2}, \ldots, y_{iq})$, where

$$y_{is} = \begin{cases} 1, & \text{if } x_i \text{ is associated with label } l_s, \\ 0, & \text{otherwise.} \end{cases}$$

Since one sample may correspond to multiple labels simultaneously, the label vector Y_i can describe the multi-label information of the sample in the decision space.

2.2 Fuzzy similarity relations

Rough set theory usually constructs approximation spaces by means of relations among objects. For interval-valued multi-label data, however, it is difficult to fully characterize inter-sample relationships by considering only conditional attributes or only label information. Therefore, fuzzy similarity relations will be defined separately in the conditional space and the decision space, based on which the corresponding granular representations are established.

Let $B \subseteq AT$ be an attribute subset. For any $x_i, x_j \in U$, let $R_B^C : U \times U \to [0,1]$ denote the binary relation induced by attribute subset B in the conditional space, where $R_B^C(x_i, x_j)$ represents the degree of similarity between samples x_i and x_j with respect to B.

Definition 2.2 [22] Let R_B^C be a binary relation induced by attribute subset B on the universe U. If, for any $x_i, x_j, x_k \in U$, R_B^C satisfies:

1. Reflexivity: $R_B^C(x_i, x_i) = 1$;
2. Symmetry: $R_B^C(x_i, x_j) = R_B^C(x_j, x_i)$;

then R_B^C is called a fuzzy similarity relation on attribute subset B. If it further satisfies

$$R_B^C(x_i, x_k) \geq \min\{R_B^C(x_i, x_j), R_B^C(x_j, x_k)\},$$

then R_B^C is called a fuzzy equivalence relation.

From the above definition, it can be seen that a fuzzy similarity relation reflects the degree of similarity between samples. Reflexivity indicates that each sample is completely similar to itself, and symmetry means that the similarity between two samples is independent of the comparison direction. If transitivity is further satisfied, a relatively stable fuzzy granular structure can be induced over the universe.

Within an interval-valued multi-label decision system, the fuzzy similarity relation defined on the conditional space measures how close samples are under interval-valued attributes, whereas the one in the decision space reflects their semantic relatedness according to label information. Based on these relations, a bi-granulation fuzzy rough set model will be further constructed, and the corresponding fuzzy positive region, dependency degree, and attribute significance will be defined.

3 An improved bi-granulation fuzzy rough set model

Based on Section 2, an improved bi-granulation fuzzy rough set model is developed in this section. Bi-granulation means that the granulation relations in both the conditional space and the decision space are simultaneously considered. For interval-valued multi-label data, however, traditional bi-granulation models still suffer from some limitations. In the conditional space, it is difficult to simultaneously characterize the midpoint position and the internal dispersion of intervals. In the decision space, the role of low-frequency labels is also not easy to reflect. Motivated by these observations, the conditional granulation relation and the decision granulation relation are respectively defined below, based on which a fuzzy rough set model with dual-space fusion is established.

3.1 Conditional granulation relation based on a center–dispersion composite interval distance

For interval-valued attributes, the difference between samples is reflected not only in the center positions of intervals but also in their internal dispersion degrees. If sample differences are described only by interval midpoints or interval boundaries, the structural characteristics of interval-valued data are often difficult to capture adequately. Inspired by the entropy-based uncertainty measurement idea in multi-granularity formal contexts[23], together with the midpoint-radius representation of intervals[24] and the integral idea of continuous interval mean absolute deviation[25], this paper constructs a composite interval distance by integrating positional difference and dispersion difference, and then defines the granulation relation in the conditional space accordingly. To eliminate the influence of attribute scales on distance calculation, all interval-valued conditional attributes are assumed to be normalized into $[0,1]$ by default.

Definition 3.1 Let the value of sample $x_i \in U$ on attribute $a \in AT$ be $a(x_i) = [a^-(x_i), a^+(x_i)]$. Then, the internal mean absolute deviation of interval $a(x_i)$ is defined as

$$\mathrm{MAD}(a(x_i)) = \frac{1}{a^+(x_i) - a^-(x_i)} \times \int_{a^-(x_i)}^{a^+(x_i)} |t - \mathrm{mid}(a(x_i))|\, dt.$$

When the internal values of the interval are approximately uniformly distributed, one has

$$\mathrm{MAD}(a(x_i)) = \frac{a^+(x_i) - a^-(x_i)}{4} = \frac{\mathrm{len}(a(x_i))}{4}.$$

Definition 3.2 Let $a(x_i)$ and $a(x_j)$ be the values of samples $x_i, x_j \in U$ on attribute $a \in AT$, respectively. The interval midpoint-position difference between them is defined as

$$d_a^{\mathrm{mid}}(x_i, x_j) = |\mathrm{mid}(a(x_i)) - \mathrm{mid}(a(x_j))|\,.$$

Definition 3.3 Let $a(x_i)$ and $a(x_j)$ be the values of samples $x_i, x_j \in U$ on attribute $a \in AT$, respectively. The internal dispersion difference between them is defined as

$$d_a^{\mathrm{mad}}(x_i, x_j) = |\mathrm{MAD}(a(x_i)) - \mathrm{MAD}(a(x_j))|\,.$$

Definition 3.4 For any samples $x_i, x_j \in U$ and attribute $a \in AT$, the composite interval distance on attribute a is defined as

$$d_a(x_i, x_j) = \alpha d_a^{\mathrm{mid}}(x_i, x_j) + (1-\alpha) d_a^{\mathrm{mad}}(x_i, x_j),$$

where $\alpha \in [0,1]$ is a balancing parameter.

From Definition 3.2 to Definition 3.4, it can be seen that $d_a^{\mathrm{mid}}(x_i, x_j)$ represents the difference in interval center positions, while $d_a^{\mathrm{mad}}(x_i, x_j)$ describes the difference in internal interval dispersion. The composite interval distance $d_a(x_i, x_j)$ integrates the above two aspects in a weighted manner. Under the uniform distribution assumption, $d_a^{\mathrm{mad}}(x_i, x_j)$ is equivalent to one fourth of the interval-length difference, and can therefore also be interpreted as the difference in uncertainty magnitude between objects.

The basic properties of the composite interval distance are given below.

Property 3.1 For any attribute $a \in AT$ and samples $x_i, x_j, x_k \in U$, $d_a(x_i, x_j)$ satisfies:

1. Non-negativity: $d_a(x_i, x_j) \geq 0$;
2. Reflexivity: $d_a(x_i, x_i) = 0$;
3. Symmetry: $d_a(x_i, x_j) = d_a(x_j, x_i)$;
4. Triangle inequality: $d_a(x_i, x_k) \leq d_a(x_i, x_j) + d_a(x_j, x_k)$.

When $0 < \alpha < 1$ and the internal values of intervals are approximately uniformly distributed,

$d_a(x_i, x_j) = 0$ if and only if $a(x_i) = a(x_j)$. Hence, d_a is a metric on the interval-valued space of attribute a.

Proof. Since both $d_a^{\mathrm{mid}}(x_i, x_j)$ and $d_a^{\mathrm{mad}}(x_i, x_j)$ are in the form of absolute values, they both satisfy non-negativity and symmetry, and both equal 0 when $i = j$. Therefore, $d_a(x_i, x_j)$ satisfies non-negativity, reflexivity, and symmetry.

By the triangle inequality of absolute values, one has

$$d_a^{\mathrm{mid}}(x_i, x_k) \le d_a^{\mathrm{mid}}(x_i, x_j) + d_a^{\mathrm{mid}}(x_j, x_k),$$
$$d_a^{\mathrm{mad}}(x_i, x_k) \le d_a^{\mathrm{mad}}(x_i, x_j) + d_a^{\mathrm{mad}}(x_j, x_k).$$

Multiplying the above inequalities by α and $1 - \alpha$, respectively, and then adding them together yields

$$d_a(x_i, x_k) \le d_a(x_i, x_j) + d_a(x_j, x_k).$$

Now assume that $0 < \alpha < 1$ and the internal values of intervals are approximately uniformly distributed. If $d_a(x_i, x_j) = 0$, then $d_a^{\mathrm{mid}}(x_i, x_j) = 0$ and $d_a^{\mathrm{mad}}(x_i, x_j) = 0$. Hence, $\mathrm{mid}(a(x_i)) = \mathrm{mid}(a(x_j))$ and $\mathrm{len}(a(x_i)) = \mathrm{len}(a(x_j))$, which implies $a(x_i) = a(x_j)$. The converse is obvious. Therefore, d_a is a metric.

Under normalization, $\mathrm{mid}(a(x_i)) \in [0, 1]$, and $\mathrm{MAD}(a(x_i)) = \mathrm{len}(a(x_i))/4 \in [0, 1/4]$. Therefore, $d_a^{\mathrm{mid}}(x_i, x_j) \in [0, 1]$ and $d_a^{\mathrm{mad}}(x_i, x_j) \in [0, 1/4]$. It follows that

$$0 \le d_a(x_i, x_j) \le \alpha + \frac{1-\alpha}{4} = \frac{1+3\alpha}{4} \le 1.$$

Moreover,

$$\frac{\partial d_a(x_i, x_j)}{\partial \alpha} = d_a^{\mathrm{mid}}(x_i, x_j) - d_a^{\mathrm{mad}}(x_i, x_j).$$

Hence, the monotonicity of $d_a(x_i, x_j)$ with respect to α depends on the relative magnitudes of $d_a^{\mathrm{mid}}(x_i, x_j)$ and $d_a^{\mathrm{mad}}(x_i, x_j)$.

Fig. 1 illustrates the construction of the composite interval distance. Specifically, the relative positions between the circle centers correspond to the midpoint-position difference $d_a^{\mathrm{mid}}(x_i, x_j)$, while the dispersion of the illustrative points within each circle and the corresponding MAD_i and MAD_j represent the internal dispersion difference $d_a^{\mathrm{mad}}(x_i, x_j)$. As shown in the figure, simultaneously considering positional difference and dispersion difference can more comprehensively characterize the discrepancy between interval-valued objects.

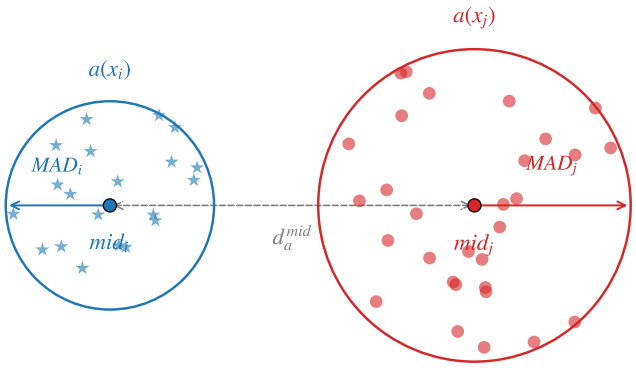


Figure 1 Illustration of the composite interval distance: center-position and internal-dispersion differences.

Definition 3.5 For any samples $x_i, x_j \in U$ and attribute $a \in AT$, the single-attribute fuzzy similarity relation in the conditional space is defined as

$$R_a^C(x_i, x_j) = \exp\big(-d_a(x_i, x_j)\big).$$

Here, $R_a^C(x_i, x_j) \in (0, 1]$, and a smaller $d_a(x_i, x_j)$ leads to a larger $R_a^C(x_i, x_j)$.

Definition 3.6 Let $B \subseteq AT$ and $B \neq \varnothing$. Then, the aggregated conditional fuzzy similarity relation between samples x_i and x_j with respect to attribute subset B is defined as

$$R_B^C(x_i, x_j) = \frac{1}{|B|} \sum_{a \in B} R_a^C(x_i, x_j).$$

The above definition adopts an averaging aggregation form to combine the single-attribute

similarities over all attributes in B. Denote

$$M_B^C = \left(R_B^C(x_i, x_j)\right)_{n\times n},$$

which is called the conditional granulation matrix corresponding to attribute subset B. Since the above definition is based on mean aggregation, $B \neq \varnothing$ is assumed by default. When a reduction algorithm is initialized with the empty set, the first-round attribute selection can be completed through the initial significance of candidate attributes, without explicitly defining $R_\varnothing^C(x_i, x_j)$.

3.2 Decision granulation relation with low-frequency enhancement

In multi-label learning, different labels often exhibit different occurrence frequencies in a dataset. High-frequency labels generally have weaker discriminative ability, whereas low-frequency labels, although appearing less often, may carry more discriminative information. If equal weights are assigned to all labels in the decision space, the contribution of low-frequency labels to sample discrimination will be weakened. Therefore, following the idea of inverse-frequency weighting, different weights are assigned to different labels, based on which the granulation relation in the decision space is defined.

Definition 3.7 Let $L = \{l_1, l_2, \ldots, l_q\}$ be the label set. For any label $l_k \in L$, its global occurrence frequency over the sample universe U is defined as

$$p_k = \frac{1}{|U|}\sum_{i=1}^{|U|} y_{ik}, \qquad k = 1, 2, \ldots, q.$$

Here, p_k represents the relative occurrence frequency of label l_k among all samples. In general, the smaller p_k is, the rarer label l_k is in the dataset, and the stronger its discriminative role tends to be.

Definition 3.8 For any label $l_k \in L$, its normalized inverse-frequency weight is defined as

$$w_k = \frac{1 - p_k}{\sum_{t=1}^{q}(1 - p_t)}, \qquad k = 1, 2, \ldots, q.$$

As long as there exists at least one label satisfying $p_k < 1$, one has $\sum_{t=1}^{q}(1-p_t) > 0$, and thus w_k is well defined. Moreover, it satisfies $w_k \geq 0$ and $\sum_{k=1}^{q} w_k = 1$. When a label has a lower global occurrence frequency, it is assigned a larger weight, so that it plays a stronger constraining role in the calculation of decision similarity.

Definition 3.9 For any samples $x_i, x_j \in U$, the weighted fuzzy similarity relation in the decision space is defined as

$$R^L(x_i, x_j) = 1 - \sum_{k=1}^{q} w_k |y_{ik} - y_{jk}|.$$

In essence, $R^L(x_i, x_j)$ is a weighted Hamming similarity based on inverse label-frequency weights, where $|y_{ik} - y_{jk}|$ is used to characterize the difference between samples x_i and x_j on label l_k. When two samples differ on low-frequency labels with large weights, their decision similarity decreases more rapidly, which is beneficial for enhancing the discriminative ability of the decision space.

Property 3.2 For any samples $x_i, x_j \in U$, the decision fuzzy similarity relation $R^L(x_i, x_j)$ satisfies:

1. Range property: $0 \leq R^L(x_i, x_j) \leq 1$;
2. Reflexivity: $R^L(x_i, x_i) = 1$;
3. Symmetry: $R^L(x_i, x_j) = R^L(x_j, x_i)$.

Proof. Since $y_{ik} \in \{0, 1\}$, one has $|y_{ik} - y_{jk}| \in \{0, 1\}$ for any k. In addition, from $w_k \geq 0$ and

$\sum_{k=1}^{q} w_k = 1$, it follows that

$$0 \le \sum_{k=1}^{q} w_k |y_{ik} - y_{jk}| \le 1.$$

Therefore,

$$0 \le R^L(x_i, x_j) = 1 - \sum_{k=1}^{q} w_k |y_{ik} - y_{jk}| \le 1.$$

When $i = j$, one has $|y_{ik} - y_{jk}| = 0$ for any k, and thus $R^L(x_i, x_i) = 1$. Moreover, since $|y_{ik} - y_{jk}| = |y_{jk} - y_{ik}|$, it follows that $R^L(x_i, x_j) = R^L(x_j, x_i)$. The proof is completed.

When two samples have exactly the same values on all labels, one has $R^L(x_i, x_j) = 1$. When they differ on low-frequency labels with relatively large weights, $R^L(x_i, x_j)$ decreases accordingly. Denote

$$M_L = \left(R^L(x_i, x_j)\right)_{n \times n},$$

which is called the decision granulation matrix induced by the decision fuzzy similarity relation R^L.

3.3 Bi-granulation fused fuzzy rough sets

For interval-valued multi-label data, the conditional space and the decision space provide two types of complementary information. The conditional space describes the structural similarity of samples with respect to interval-valued attributes, whereas the decision space characterizes the supervisory relationships among samples in terms of labels. Based on the conditional granulation relation and the decision granulation relation, a dual-space fused fuzzy relation is introduced below, upon which a bi-granulation fused fuzzy rough set model is established.

3.3.1 Single-parameter dual-space fused granulation similarity relation

Definition 3.10 Let $B \subseteq AT$ and $B \neq \varnothing$. For any samples $x_i, x_j \in U$, the dual-space fused fuzzy similarity relation with respect to attribute subset B is defined as

$$R_B^{CL}(x_i, x_j) = \beta R_B^C(x_i, x_j) + (1 - \beta) R^L(x_i, x_j),$$

where $\beta \in [0, 1]$ is a balancing parameter between the two spaces.

In Definition 3.10, the parameter β adjusts the relative contributions of the conditional space and the decision space in the fused similarity relation. When β is large, the fused relation places more emphasis on the structural information in the conditional space; when β is small, it relies more on the label supervisory information in the decision space. Denote

$$M_B^{CL} = \left(R_B^{CL}(x_i, x_j)\right)_{n \times n},$$

which is called the dual-space fused granulation matrix corresponding to attribute subset B.

Property 3.3 For any $B \subseteq AT$ with $B \neq \varnothing$, and any samples $x_i, x_j \in U$, the dual-space fused fuzzy similarity relation $R_B^{CL}(x_i, x_j)$ satisfies:

1. $0 \le R_B^{CL}(x_i, x_j) \le 1$;
2. $R_B^{CL}(x_i, x_i) = 1$;
3. $R_B^{CL}(x_i, x_j) = R_B^{CL}(x_j, x_i)$.

Proof. By the definitions of the fuzzy similarity relation in the conditional space and that in the decision space, for any $x_i, x_j \in U$, one has $R_B^C(x_i, x_j) \in (0, 1]$, $R^L(x_i, x_j) \in [0, 1]$, and $\beta \in [0, 1]$. Therefore, $R_B^{CL}(x_i, x_j)$ is a convex combination of $R_B^C(x_i, x_j)$ and $R^L(x_i, x_j)$, which implies $0 \le R_B^{CL}(x_i, x_j) \le 1$.

Moreover, since $R_B^C(x_i, x_i) = 1$ and $R^L(x_i, x_i) = 1$, one obtains

$$R_B^{CL}(x_i, x_i) = \beta \cdot 1 + (1 - \beta) \cdot 1 = 1.$$

Further, from $R_B^C(x_i, x_j) = R_B^C(x_j, x_i)$ and

$R^L(x_i, x_j) = R^L(x_j, x_i)$, it follows that

$$R_B^{CL}(x_i, x_j) = R_B^{CL}(x_j, x_i).$$

The proof is completed.

3.3.2 Fused fuzzy positive region and dependency degree

Using the dual-space fused granulation relation, we next formulate the fused fuzzy positive region for the label set L together with its associated dependency degree.

Definition 3.11 Let $B \subseteq AT$ and $B \neq \varnothing$. For each $x_i \in U$, the membership of x_i in the fused fuzzy positive region associated with the label set L is given by

$$\mu_{\mathrm{POS}_B(L)}(x_i) = \frac{1}{|U|} \sum_{x_j \in U} R_B^{CL}(x_i, x_j) R^L(x_i, x_j).$$

Accordingly, The fused fuzzy positive region of the attribute subset B for L is then written as

$$\mathrm{POS}_B(L) = \left\{ \frac{\mu_{\mathrm{POS}_B(L)}(x_i)}{x_i} \,\middle|\, x_i \in U \right\},$$

and its fuzzy cardinality is defined as

$$|\mathrm{POS}_B(L)| = \sum_{x_i \in U} \mu_{\mathrm{POS}_B(L)}(x_i).$$

The dependency degree of B on the label set L is expressed as

$$\gamma_B(L) = \frac{|\mathrm{POS}_B(L)|}{|U|}.$$

In Definition 3.11, $\mu_{\mathrm{POS}_B(L)}(x_i)$ refers to the membership value of sample x_i in the fused fuzzy positive region. The notation $\mathrm{POS}_B(L)$ stands for the fused fuzzy positive region generated by the attribute subset B for the label set L, and $|\mathrm{POS}_B(L)|$ denotes its fuzzy cardinality. Moreover, $\gamma_B(L)$ measures the dependency of B on L. This formulation evaluates the degree to which a sample belongs to the positive region by jointly considering the fused granulation relation R_B^{CL} and the decision similarity relation R^L.

Property 3.4 For any $B \subseteq AT$ with $B \neq \varnothing$, one has

$$0 \leq \mu_{\mathrm{POS}_B(L)}(x_i) \leq 1, \qquad 0 \leq \gamma_B(L) \leq 1.$$

Proof. According to Property 3.3, one has $R_B^{CL}(x_i, x_j) \in [0, 1]$ for any $x_i, x_j \in U$. Moreover, by the definition of the fuzzy similarity relation in the decision space, $R^L(x_i, x_j) \in [0, 1]$. Therefore,

$$0 \leq R_B^{CL}(x_i, x_j) R^L(x_i, x_j) \leq 1.$$

Taking the average over $x_j \in U$, one obtains

$$0 \leq \mu_{\mathrm{POS}_B(L)}(x_i) \leq 1.$$

Furthermore,

$$0 \leq |\mathrm{POS}_B(L)| = \sum_{x_i \in U} \mu_{\mathrm{POS}_B(L)}(x_i) \leq |U|,$$

which yields

$$0 \leq \gamma_B(L) = \frac{|\mathrm{POS}_B(L)|}{|U|} \leq 1.$$

The proof is completed.

A larger value of $\gamma_B(L)$ indicates that attribute subset B preserves the label information to a greater extent. It should be noted that the conditional-space similarity relation R_B^C is obtained by averaging the single-attribute similarities over the attribute subset. Therefore, $\gamma_B(L)$ does not necessarily satisfy strict monotonicity in general. Accordingly, $\gamma_B(L)$ is mainly regarded here as an attribute evaluation criterion under the bi-granulation fused fuzzy rough model.

3.3.3 Attribute significance

The significance of a candidate attribute is evaluated by its additional contribution to the preservation capability of the current attribute subset.

Definition 3.12 For any $B \subseteq AT$ and $a \in AT \setminus B$, the significance of attribute a relative to attribute subset B is defined as

$$\mathrm{Sig}(a, B, L) = \gamma_{B\cup\{a\}}(L) - \gamma_B(L).$$

Definition 3.12 represents the change in dependency degree after adding attribute a to attribute subset B. The larger $\mathrm{Sig}(a, B, L)$ is, the greater the gain brought by attribute a to the current attribute subset. Since the dependency degree $\gamma_B(L)$ does not generally satisfy strict monotonicity, $\mathrm{Sig}(a, B, L)$ may be positive or negative. When $\mathrm{Sig}(a, B, L) \leq 0$, attribute a brings no positive gain, and therefore need not be given priority in the subsequent reduction process.

4 Feature selection method based on the improved bi-granulation fuzzy rough set model

4.1 A forward greedy feature selection algorithm

The dependency degree $\gamma_B(L)$ is used to evaluate attribute subsets, and the attribute significance $\mathrm{Sig}(a, B, L)$ is adopted to rank candidate attributes. Based on these two measures, a forward greedy strategy is employed to construct a reduct incrementally.

Since the aggregated conditional fuzzy similarity relation R_B^C in Definition 3.6 is defined only for nonempty subsets B, the search starts from single attributes. The attribute with the largest dependency degree is selected as the initial reduct element. Then, at each step, the remaining attribute with the greatest significance is added. Because mean aggregation does not guarantee strict monotonicity of the dependency degree with respect to attribute inclusion, the stopping condition is determined by the marginal gain relative to the current reduct. The search terminates when the maximum gain is smaller than a given threshold ε.

Algorithm 1: IVMDG-F: A forward greedy feature selection algorithm

Input: $IV\text{-}MLDS = \langle U, AT, L\rangle$, parameters α, β, threshold ε
Output: Reduct Red
foreach $a \in AT$ **do**
  Compute $\gamma_{\{a\}}(L)$
end
Select the attribute a^* with the maximum dependency degree
$a^* \leftarrow \arg\max_{a\in AT} \gamma_{\{a\}}(L)$
$Red \leftarrow \{a^*\}$
while $AT \setminus Red \neq \varnothing$ **do**
  foreach $a \in AT \setminus Red$ **do**
    Compute $\gamma_{Red\cup\{a\}}(L)$
    $\mathrm{Sig}(a, Red, L) \leftarrow \gamma_{Red\cup\{a\}}(L) - \gamma_{Red}(L)$
  end
  Select the attribute a^* with the maximum significance
  $a^* \leftarrow \arg\max_{a\in AT\setminus Red} \mathrm{Sig}(a, Red, L)$
  if $\mathrm{Sig}(a^*, Red, L) \geq \varepsilon$ **then**
    $Red \leftarrow Red \cup \{a^*\}$
  else
    return Red
  end
end
return Red

The corresponding procedure is given in Algorithm 1.

For complexity analysis, let $n = |U|$, $m = |AT|$, and $q = |L|$ denote the numbers of samples, attributes, and labels, respectively. Constructing the weighted fuzzy similarity relation in the decision space requires $O(n^2q)$ time. If

the conditional similarity matrices of all individual attributes are precomputed and stored, evaluating one candidate attribute requires $O(n^2)$ time. In the worst case, the forward search performs $O(m^2)$ such evaluations, yielding an overall time complexity of $O(m^2n^2+n^2q)$. Moreover, storing all single-attribute conditional similarity matrices together with the decision-space similarity matrix requires $O(mn^2)$ space.

4.2 Illustrative example

To illustrate the computational procedure of the proposed method, we construct an interval-valued multi-label decision system $IV\text{-}MLDS = \langle U, AT, L\rangle$, where $U = \{x_1, x_2, x_3, x_4\}$, $AT = \{a_1, a_2, a_3\}$, and $L = \{l_1, l_2, l_3\}$. The system is shown in Table 1.

Table 1 An example of an interval-valued multi-label decision system

U	a_1	a_2	a_3	l_1	l_2	l_3
x_1	$[0.20, 0.40]$	$[0.60, 0.80]$	$[0.30, 0.50]$	1	1	0
x_2	$[0.30, 0.50]$	$[0.55, 0.75]$	$[0.35, 0.55]$	1	0	0
x_3	$[0.70, 0.90]$	$[0.20, 0.40]$	$[0.60, 0.80]$	0	1	1
x_4	$[0.65, 0.85]$	$[0.25, 0.45]$	$[0.58, 0.78]$	0	1	0

Let the parameters be $\alpha = 0.5$, $\beta = 0.5$, and $\varepsilon = 0.02$.

First, according to Definitions 3.1–3.5, the conditional fuzzy similarity relation on each attribute is calculated. Taking samples x_1 and x_2 under attribute a_1 as an example, we have

$$\begin{aligned}&\mathrm{mid}(a_1(x_1)) = 0.30, \quad \mathrm{mid}(a_1(x_2)) = 0.40,\\ &\mathrm{MAD}(a_1(x_1)) = \mathrm{MAD}(a_1(x_2)) = 0.05.\end{aligned}$$

Thus,

$$\begin{aligned}&d_{a_1}^{\mathrm{mid}}(x_1, x_2) = 0.10, \quad d_{a_1}^{\mathrm{mad}}(x_1, x_2) = 0,\\ &d_{a_1}(x_1, x_2) = 0.5 \times 0.10 + 0.5 \times 0 = 0.05,\\ &R_{a_1}^{C}(x_1, x_2) = \exp(-0.05) = 0.9512.\end{aligned}$$

Similarly, the conditional fuzzy similarity matrices for all single attributes can be obtained. For attribute a_1, the matrix is

$$M_{a_1} = \begin{pmatrix} 1.0000 & 0.9512 & 0.7788 & 0.7985 \\ 0.9512 & 1.0000 & 0.8187 & 0.8395 \\ 0.7788 & 0.8187 & 1.0000 & 0.9753 \\ 0.7985 & 0.8395 & 0.9753 & 1.0000 \end{pmatrix}.$$

The fuzzy similarity relation in the decision space is derived from Definitions 3.7–3.9. From the label distribution, we obtain $p_1 = 0.5$, $p_2 = 0.75$, and $p_3 = 0.25$. The corresponding normalized inverse-frequency weights are

$$w_1 = \frac{1}{3}, \quad w_2 = \frac{1}{6}, \quad w_3 = \frac{1}{2}.$$

Hence, the decision granulation matrix is

$$M_L = \begin{pmatrix} 1.0000 & 0.8333 & 0.1667 & 0.6667 \\ 0.8333 & 1.0000 & 0.0000 & 0.5000 \\ 0.1667 & 0.0000 & 1.0000 & 0.5000 \\ 0.6667 & 0.5000 & 0.5000 & 1.0000 \end{pmatrix}.$$

Based on this relation, Definitions 3.10 and 3.11 are applied to evaluate the dependency degree of each attribute subset with respect to the label set L. Taking the attribute subset $\{a_1\}$ as an example, we obtain $\gamma_{\{a_1\}}(L) = 0.5018$. Similarly, the dependency degrees of all attribute subsets are listed in Table 2.

Table 2 Dependency degrees of different attribute subsets

Attribute subset	Dependency degree
$\{a_1\}$	0.5018
$\{a_2\}$	0.4992
$\{a_3\}$	0.5033
$\{a_1, a_2\}$	0.5005
$\{a_1, a_3\}$	0.5025
$\{a_2, a_3\}$	0.5012
$\{a_1, a_2, a_3\}$	0.5014

Finally, attribute reduction is performed.

From Table 2, it can be seen that among all single-attribute subsets, $\gamma_{\{a_3\}}(L) = 0.5033$ is the largest. Therefore, the initial reduct is $Red = \{a_3\}$.

For the remaining attributes a_1 and a_2, their significance with respect to the current reduct Red is calculated as follows:

$$\begin{aligned}\mathrm{Sig}(a_1, \{a_3\}, L) &= \gamma_{\{a_1,a_3\}}(L) - \gamma_{\{a_3\}}(L) \\ &= 0.5025 - 0.5033 = -0.0008, \\ \mathrm{Sig}(a_2, \{a_3\}, L) &= \gamma_{\{a_2,a_3\}}(L) - \gamma_{\{a_3\}}(L) \\ &= 0.5012 - 0.5033 = -0.0021.\end{aligned}$$

Therefore,

$$\max\big\{\mathrm{Sig}(a_1, \{a_3\}, L), \mathrm{Sig}(a_2, \{a_3\}, L)\big\} = -0.0008 < \varepsilon.$$

This indicates that adding any of the remaining attributes cannot produce a gain no less than the threshold ε. Therefore, the algorithm stops, and the final reduct is $Red = \{a_3\}$. Since the defined dependency degree is generally not strictly monotonic, the attribute reduction process adopts a stopping criterion based on marginal gain.

5 Experimental results and analysis

5.1 Description of experimental datasets

Eight real-world multi-label datasets were used in the experiments. Table 3 summarizes the numbers of original features, selected features, and the corresponding feature retention rates. These datasets differ substantially in feature dimensionality, covering both low-dimensional and high-dimensional cases, and thus provide a suitable basis for testing the proposed method under different feature scales.

Table 3 Selected features and retention rates on eight datasets

Dataset	Original features	Selected features	Rate (%)
Emotions	72	22	30.56
Flags	19	4	21.05
Birds	260	78	30.00
CAL500	68	14	20.59
Image	294	60	20.41
Scene	294	57	19.39
Yeast	103	22	21.36
Education	500	132	26.40

5.2 Experimental settings and result analysis

The proposed method was compared with several representative multi-label feature selection approaches, namely DGFRS [26], NRFSFN [27], ENFSFN [27], and MLDRS [28].

Table 4 summarizes the results on the eight datasets with respect to five evaluation metrics. In the table, "↓" denotes that lower values are preferred, while "↑" denotes that higher values are preferred. The best value under each metric is highlighted in bold.

As shown in Table 4, IVMDG-F exhibits competitive overall performance across the eight datasets and five evaluation metrics. This suggests that the proposed method can retain important information for multi-label classification after feature reduction. More specifically, IVMDG-F achieves the best performance on six datasets for Hamming Loss, Coverage, and Average Precision, and on five datasets for Ranking Loss and One-Error. These findings indicate that the proposed method is effective in lowering misclassification, improving label ranking quality, and identifying relevant labels.

From the perspective of average rank, IVMDG-F achieves average ranks of 1.75, 2.00, 1.38, 1.63, and 1.88 under the five evaluation metrics, respectively, with an overall average

Table 4 Comparison of evaluation metrics on eight datasets

Metric	Dataset	DGFRS	NRFSFN	ENFSFN	MLDRS	IVMDG-F
	Emotions	0.1990 2	0.2074 4	0.1998 3	0.2119 5	**0.1937 1**
	Flags	0.3075 4	0.2959 2.5	0.2959 2.5	0.3260 5	**0.2939 1**
	Birds	0.0563 5	0.0515 3	0.0512 2	**0.0510 1**	0.0562 4
	CAL500	0.1402 5	0.1381 2	0.1386 3	0.1389 4	**0.1332 1**
HL↓	Education	0.0438 5	0.0384 3	0.0383 2	**0.0336 1**	0.0395 4
	Image	0.2075 4	0.1893 2	0.1931 3	0.2078 5	**0.1867 1**
	Scene	0.1153 4	0.1105 3	0.1191 5	0.1059 2	**0.1037 1**
	Yeast	0.1975 3	0.1989 4	0.2024 5	0.1799 2	**0.1785 1**
	Average rank	4.00	2.94	3.19	3.13	**1.75**
	Emotions	0.1648 2	0.1708 4	0.1682 3	0.1865 5	**0.1593 1**
	Flags	0.2236 3	0.2363 4.5	0.2363 4.5	0.2149 2	**0.2106 1**
	Birds	0.2190 4	0.2119 2	**0.2113 1**	0.2285 5	0.2146 3
	CAL500	0.1872 5	0.1829 2	0.1831 3	0.1836 4	**0.1827 1**
RL↓	Education	0.0864 3	0.0824 2	**0.0742 1**	0.0938 5	0.0877 4
	Image	0.1985 3	0.1973 2	0.1992 4	0.2352 5	**0.1854 1**
	Scene	0.1686 5	0.1197 3	0.1170 2	0.1488 4	**0.1169 1**
	Yeast	0.1781 3	0.1760 2	0.1786 5	**0.1469 1**	0.1783 4
	Average rank	3.50	2.69	2.94	3.88	**2.00**
	Emotions	1.8009 2	1.8317 4	1.8179 3	1.9193 5	**1.7949 1**
	Flags	3.9481 5	3.9367 3	3.9397 4	3.9233 2	**3.9197 1**
	Birds	3.0642 3	5.8478 4	5.8960 5	**2.8192 1**	2.9227 2
	CAL500	130.8643 5	130.4577 4	130.2556 3	129.9829 2	**129.9359 1**
CV↓	Education	3.9446 4	3.5525 2	**3.4999 1**	4.1267 5	3.8233 3
	Image	1.1602 4	1.0580 2	1.0650 3	1.1902 5	**1.0250 1**
	Scene	0.8564 5	0.6353 2	0.6722 3	0.8217 4	**0.6334 1**
	Yeast	6.4942 5	6.4506 4	6.4303 2	6.4499 3	**6.4051 1**
	Average rank	4.13	3.13	3.00	3.38	**1.38**
	Emotions	0.8005 2	0.7898 4	0.7927 3	0.7893 5	**0.8098 1**
	Flags	0.8063 3	0.8021 4.5	0.8021 4.5	**0.8236 1**	0.8132 2
	Birds	0.5274 4	0.5287 3	0.5355 2	0.5121 5	**0.5369 1**
	CAL500	0.4920 4	0.4925 3	0.4929 2	0.4794 5	**0.4961 1**
AP↑	Education	0.5697 5	0.5906 3	0.6241 2	**0.6406 1**	0.5872 4
	Image	**0.7952 1**	0.7655 3	0.7643 4	0.7396 5	0.7897 2
	Scene	0.8195 2	0.8169 3	0.8014 4	0.7616 5	**0.8418 1**
	Yeast	0.7559 3	0.7543 4	0.7477 5	0.7792 2	**0.7871 1**
	Average rank	3.00	3.44	3.31	3.63	**1.63**
	Emotions	0.2427 2	0.2900 4	0.2866 3	0.3349 5	**0.2421 1**
	Flags	0.2202 5	**0.2112 1.5**	**0.2112 1.5**	0.2193 4	0.2165 3
	Birds	**0.5398 1**	0.7550 5	0.7473 4	0.5669 2	0.5845 3
	CAL500	0.1185 5	0.1156 3.5	0.1156 3.5	0.1103 2	**0.1093 1**
OE↓	Education	0.5735 4	0.5387 3	0.5243 2	0.6228 5	**0.5232 1**
	Image	**0.3506 1**	0.3610 2	0.3625 3	0.4184 5	0.3975 4
	Scene	0.3433 4	0.3170 2	0.3340 3	0.3772 5	**0.3168 1**
	Yeast	0.2359 4	0.2333 3	0.2445 5	0.2238 2	**0.2209 1**
	Average rank	3.25	3.00	3.13	3.75	**1.88**
Overall rank over five metrics		5	2	3	4	**1**
Average ranking value over five metrics		3.58	3.04	3.11	3.55	**1.73**

rank of 1.73, ranking first among all compared methods. This indicates that the proposed method consistently performs well across different datasets and evaluation metrics. In particular, on datasets such as Image, Scene, and Yeast, IVMDG-F achieves favorable results under multiple metrics, which further demonstrates its robustness on these datasets. Although IVMDG-F does not always achieve the best result on every individual dataset, the gap between IVMDG-F and the best-performing method is generally small, and its performance remains relatively stable.

To analyze the influence of feature subset size on performance, the Yeast dataset is taken as an example, and the variation curves of the evaluation metrics are plotted in Figure 2. As the number of selected features increases, the performance of all methods gradually becomes stable. This indicates that when the feature subset size reaches a certain level, further increasing the number of features yields only limited improvement. IVMDG-F consistently shows favorable performance in terms of metrics such as AP and Coverage. In particular, it can still achieve competitive results when the number of selected features is relatively small, indicating that the proposed method is able to preserve a considerable amount of useful information even with a compact feature subset, while maintaining good stability.

5.3 Statistical significance analysis

To examine whether the performance differences among the compared algorithms are statistically significant, the Friedman test was first employed, followed by the Nemenyi post-hoc test when a significant overall difference was detected. In the Friedman test, the null hypothesis assumes that the compared algorithms do not differ significantly in performance across the datasets.

The Friedman test statistics are defined as

$$\chi_F^2 = \frac{12T}{s(s+1)} \left(\sum_{i=1}^{s} R_i^2 - \frac{s(s+1)^2}{4} \right), \quad (5.1)$$

$$F_F = \frac{(T-1)\chi_F^2}{T(s-1) - \chi_F^2}, \quad (5.2)$$

where T is the number of datasets, s is the number of compared algorithms, and R_i is the mean rank of the i-th algorithm across all datasets.

The Friedman test results for the five evaluation metrics are summarized in Table 5. In this experiment, $s = 5$ and $T = 8$. Under the significance level $\alpha = 0.05$, the critical value is $F_{0.05}(4, 28) = 2.7141$.

Table 5 Friedman statistical results for different evaluation metrics

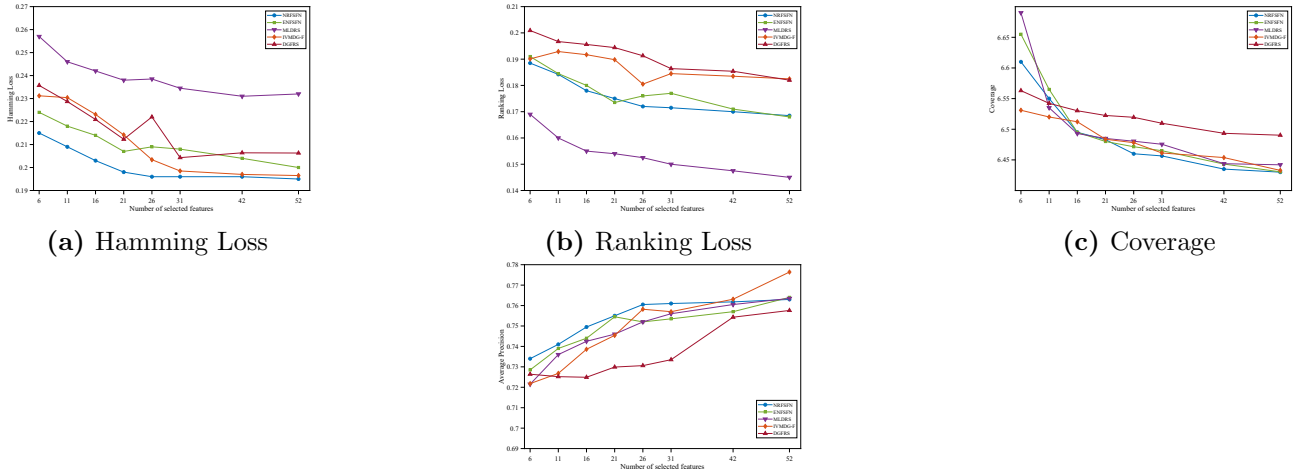
Metric	F_F	Critical value ($\alpha = 0.05$)
HL	2.5027	2.7141
RL	1.8951	
CV	4.7895	
AP	2.4422	
OE	1.6745	

According to Table 5, only the statistic for Coverage exceeds the critical value, so the null hypothesis is rejected for this metric. For the remaining four metrics, all statistics are below the critical value. Hence, the null hypothesis is not rejected, indicating that the observed performance differences among the compared algorithms are not statistically significant under these metrics.

For the Coverage metric, the Nemenyi post-hoc test was further conducted for pairwise comparisons. Its critical difference is defined as

$$CD_\alpha = q_\alpha \sqrt{\frac{s(s+1)}{6T}}, \quad (5.3)$$

where q_α is determined by the significance level α and the number of algorithms s. Under the present setting ($s = 5, T = 8$), when $\alpha = 0.05$,



(a) Hamming Loss **(b)** Ranking Loss **(c)** Coverage

(d) Average Precision

Figure 2 Changes of four metrics of different methods with different feature subset sizes on the Yeast dataset

we take $q_\alpha = 2.728$, and the critical difference is

$$CD = 2.728\sqrt{\frac{5(5+1)}{6\times 8}} = 2.1567. \qquad (5.4)$$

According to the Nemenyi test results, for the Coverage metric, IVMDG-F achieves the best average rank of 1.375, while DGFRS has an average rank of 4.125. The difference in their average ranks is 2.750, which is larger than the critical difference 2.1567. This indicates that there exists a statistically significant difference between IVMDG-F and DGFRS, and IVMDG-F performs better. In contrast, the differences between IVMDG-F and the remaining methods do not exceed the critical difference. This can also be observed more intuitively from Figure 3.

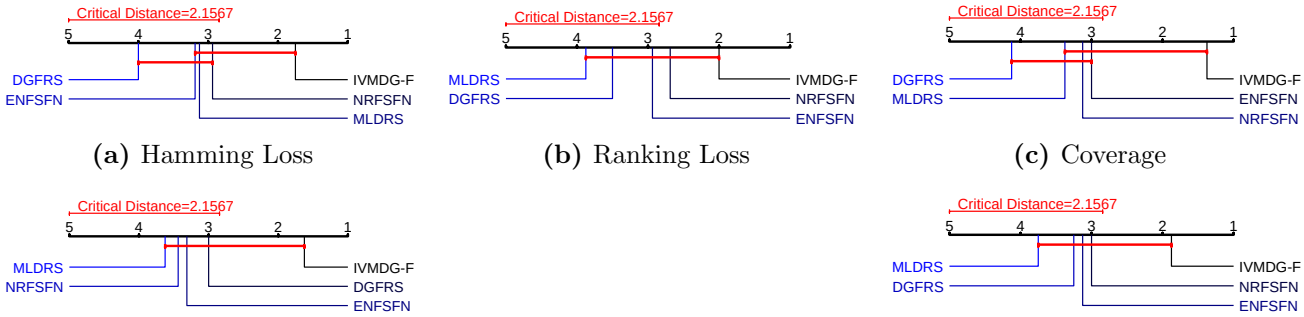
Taken together, the results of the Friedman test and the Nemenyi post-hoc test show that the proposed method has a statistically significant advantage in terms of Coverage. Although the differences among the compared methods on the other evaluation metrics do not reach statistical significance, IVMDG-F still ranks highly in terms of average rank and overall rank across most metrics, which further demonstrates the stability of the proposed method.

5.4 Parameter sensitivity analysis

In the proposed method, parameter α adjusts the balance between the positional difference and the dispersion difference in the composite interval distance, while parameter β coordinates the weights of the similarity relation in the conditional space and that in the decision space during the fusion process. Both parameters affect the calculation of inter-sample similarity and, consequently, the feature reduction results. Therefore, it is necessary to analyze their influence on the model performance.

A grid search strategy was adopted for parameter testing. Specifically, α was varied from 0.4 to 0.8 with a step size of 0.1, and β was varied from 0.80 to 1.00 with a step size of 0.05. Under the same experimental settings, Average Precision (AP) and Hamming Loss (HL) were selected as the evaluation metrics. The results are shown in Figure 4.

It can be observed from Figure 4(a) that AP fluctuates only slightly under different parameter combinations, indicating that the proposed method has good stability. When α ranges from 0.5 to 0.7, AP tends to achieve relatively favorable performance. From Figure 4(b), it can be seen that HL is more sensitive to parameter changes. As α increases, HL



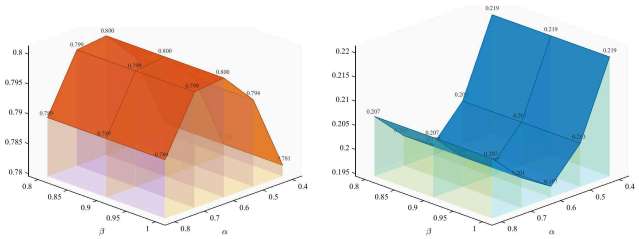
(a) Hamming Loss **(b)** Ranking Loss **(c)** Coverage

(d) Average Precision **(e)** One-Error

Figure 3 Results of the Nemenyi test

generally shows a decreasing trend, suggesting that an appropriate increase in α helps reduce classification error, while the influence of β is relatively smaller.

Considering both AP and HL, it can be concluded that the model achieves better overall performance when α takes a moderately large value. Therefore, $\alpha = 0.6$ and $\beta = 0.90$ are finally selected as the default parameter settings in the subsequent experiments.



(a) Average Precision (b) Hamming Loss

Figure 4 Effects of parameters α and β on model performance

6 Conclusion

In interval-valued multi-label decision systems, the conditional part is affected by interval uncertainty, while the decision part carries multi-label semantic information. To cope with these characteristics, this study develops an improved bi-granulation interval-valued multi-label fuzzy rough feature selection method. In the conditional space, the proposed method combines the positional difference of interval midpoints with the internal dispersion difference, while in the decision space, normalized inverse label-frequency weights are introduced. Based on this framework, the fuzzy positive region, dependency degree, and attribute significance are further defined. Experimental results show that the proposed method achieves favorable results on multiple evaluation metrics and exhibits relatively stable performance. Since the construction and fusion of sample relations are relatively complex, the computational cost remains high when the data scale becomes large. Future work will further consider improving algorithmic efficiency and modeling label correlations.

Author Contributions

Shuwen Zheng: conceptualization, writing–original draft, and methodology. Guoping Lin: writing–review and editing, supervision, and funding acquisition. Weihua Xu: validation, formal analysis, and writing–review and editing. Yujie Qin: writing–review and editing, software, and data curation.

Acknowledgments

This work was supported by the National Natural Science Foundation of China, No. 624B2104; the 2026 Open Fund Project of the Key Laboratory of Big Data Mining and Application in Fujian Universities under Grant No. DSJZDSYS202602; and the Graduate Teaching Reform Project of Minnan Normal University under Grant No. YJG202630.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that have been used are confidential.

References

[1] MUHAMMAD S, USMAN Q. Redefining core preliminary concepts of classic rough set theory for feature selection[J]. Engineering Applications of Artificial Intelligence, 2017, 65: 375–387.

[2] HU Q H, YU D R, LIU J F, WU C X. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences, 2008, 178(18): 3577–3594.

[3] PAWLAK Z. Rough sets[J]. International Journal of Computer & Information Sciences, 1982, 11(5): 341–356.

[4] DUBOIS D, PRADE H. Rough fuzzy sets and fuzzy rough sets[J]. International Journal of General Systems, 1990, 17(2/3): 191–209.

[5] WANG C Z, QI Y L, SHAO M W, et al. A fitting model for feature selection with fuzzy rough sets[J]. IEEE Transactions on Fuzzy Systems, 2017, 25(4): 741–753.

[6] HU Q H, YU D R, LIU J F, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences, 2008, 178(18): 3577–3594.

[7] WANG C Z, SHAO M W, HE Q, et al. Feature subset selection based on fuzzy neighborhood rough sets[J]. Knowledge-Based Systems, 2016, 111: 173–179.

[8] XIA S Y, ZHANG H, WANG G Y, et al. GBNRS: A novel rough set algorithm for fast adaptive attribute reduction in classification[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(3): 1231–1242.

[9] PENG X L, WANG P, YANG Y, et al. FNC: A fast neighborhood calculation framework[J]. Knowledge-Based Systems, 2022, 252: 109394.

[10] WANG C Z, WANG C Y, GAO Y, et al. Feature selection based on weighted fuzzy rough sets[J]. IEEE Transactions on Fuzzy Systems, 2024, 32(7): 4027–4037.

[11] DAI J, HUANG W, ZHANG C, et al. Multi-label feature selection by strongly relevant label gain and label mutual aid[J]. Pattern Recognition, 2024, 145: 109945.

[12] DENG Z, LI T, DENG D, et al. Feature selection for handling label ambiguity using weighted label-fuzzy relevancy and redundancy[J]. IEEE Transactions on Fuzzy Systems, 2024, 32(8): 4436–4447.

[13] ZOU Y, HU X, LI P. Gradient-based multi-label feature selection considering three-way variable interaction[J]. Pattern Recognition, 2024, 145: 109900.

[14] MOHANRAJU S S, JANANI K, KEERTHANA J, et al. Multi-scale multi-label feature selection via fuzzy mutual information[J]. International Journal of Machine Learning and Cybernetics, 2025, 16(12): 1–19.

[15] ZHANG Q L, LIU S P, WANG J, et al. Feature selection for multi-labeled data based on label enhancement technique and mutual information[J]. Information Sciences, 2024, 679: 121113.

[16] YU Y, JIANG Z Z, LU Z, et al. Dual-perspective imbalance-aware feature selection for multi-label learning[J]. European Journal of Operational Research, 2026, 335(3): 919–932.

[17] DAI J H, WU Q Y, PEDRYCZ W. Noise-resistant commonality and individuality label learning for multiview multilabel feature selection using fuzzy mutual information[J]. IEEE Transactions on Cybernetics, 2026, 56(3): 1479–1491.

[18] QIAN Y, LIANG J, DANG C. Interval ordered information systems[J]. Computers and Mathematics with Applications, 2008, 56(8): 1994–2009.

[19] LI W, ZHOU H, XU W, et al. Interval dominance-based feature selection for interval-valued ordered data[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 34(10): 6898–6912.

[20] MA X, ZHANG Q, QIU J. Semi-supervised feature selection for partially labeled interval-valued information systems[J]. Cluster Computing, 2026, 29(2): 96.

[21] BURUSCO A, FUENTES-GONZALEZ R. The study of interval-valued contexts[J]. Fuzzy Sets and Systems, 2001, 121(3): 439–452.

[22] LI Y H, HU L, GAO W F. Label correlations variation for robust multi-label feature selection[J]. Information Sciences, 2022, 609: 1075–1097.

[23] ZHANG X Y, ZHAO W C. Uncertainty measures and feature selection based on composite entropy for generalized multigranulation fuzzy neighborhood rough set[J]. Fuzzy Sets and Systems, 2024, 486: 108971.

[24] PALUMBO F, LAURO C N. A PCA for interval-valued data based on midpoints and radii[C]//New

Developments in Psychometrics. Tokyo: Springer, 2003: 641–648.

[25] PINSKY E. Computation and interpretation of mean absolute deviations by cumulative distribution functions[J]. Frontiers in Applied Mathematics and Statistics, 2025, 11: 1487331.

[26] QIN Y, LIN G, LIN Y, et al. Dominance relation-based feature selection for interval-valued multi-label ordered information system[J]. Expert Systems with Applications, 2025, 274: 126898.

[27] YIN T, CHEN H, YUAN Z, et al. Noise-resistant multilabel fuzzy neighborhood rough sets for feature subset selection[J]. Information Sciences, 2023, 621: 200–226.

[28] KOU Y, LIN G, QIAN Y, et al. A novel multi-label feature selection method with association rules and rough set[J]. Information Sciences, 2023, 624: 299–323.