Multifuzzy β-Covering Approximation Spaces and Their Information Measures

Jianhua Dai, Xiongtao Zou, Yuhua Qian, *Member, IEEE*, and Xizhao Wang, *Fellow, IEEE*

***Abstract*—Fuzzy β-covering rough sets, as an effective extension of covering-based rough sets, have been concerned by many researchers. All fuzzy β-covering rough set models are constructed under a corresponding fuzzy β-covering approximation space. However, fuzzy β-covering is difficult to find directly from the real data. Fortunately, fuzzy information granulation provides a reasonable and effective way to obtain fuzzy β-coverings from the real data. Since fuzzy information granulation is capable of generating multiple fuzzy β-coverings, we introduce the notion of multifuzzy β-covering approximation spaces. Fuzzy β-covering approximation spaces are a special case of multifuzzy β-covering approximation spaces. Besides, we employ fuzzy β-neighborhood operators with reflexivity and symmetry to characterize the similarity between samples. In this article, we first present the definition of multifuzzy β-covering approximation spaces and investigate some useful properties about fuzzy β-covering. Second, several information measures are explored in the context of multifuzzy β-covering approximation spaces. On this basis, a novel heuristic fuzzy β-covering reduction method with the measure of monotone conditional entropy is proposed. Moreover, a general framework of attribute reduction based on fuzzy β-covering reduction is also designed. Finally, through the comparative and experimental analyses with other four state-of-the-art attribute reduction methods, the effectiveness and superiority of the proposed method are verified.**

***Index Terms*—Attribute reduction, fuzzy β-covering, fuzzy rough sets, granular computing, uncertainty measure.**

Manuscript received 7 June 2022; accepted 13 July 2022. Date of publication 25 July 2022; date of current version 1 March 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 61976089, in part by the Major Program of the National Social Science Foundation of China under Grant 20&ZD047, in part by the Natural Science Foundation of Hunan Province under Grant 2021JJ30451 and Grant 2022JJ30397, in part by the Hunan Provincial Science and Technology Project Foundation under Grant 2018RS3065 and Grant 2018TP1018, and in part by the Postgraduate Scientific Research Innovation Project of Hunan Province under Grant CX20210431. *(Corresponding author: Jianhua Dai.)*

Jianhua Dai and Xiongtao Zou are with the Hunan Provincial Key Laboratory of Intelligent Computing and Language Information Processing, Hunan Normal University, Changsha 410081, China, and also with the College of Information Science and Engineering, Hunan Normal University, Changsha 410081, China (e-mail: jhdai@hunnu.edu.cn; zxt_hnnu@163.com).

Yuhua Qian is with the Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China, and also with the Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China (e-mail: jinchengqyh@126.com).

Xizhao Wang is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China, and also with the Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China (e-mail: xizhaowang@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TFUZZ.2022.3193448.

Digital Object Identifier 10.1109/TFUZZ.2022.3193448

I. Introduction

COVERING-BASED rough set model, as an effective extension of classical rough set model [1], was first proposed by Zakowski [2]. Due to its strong ability of uncertain reasoning, the theory and application of covering-based rough sets [3]–[5] have further been developed. However, these models cannot directly process the data with real values. For this reason, Dubois and Prade [6] put forward the theory of fuzzy rough sets by integrating the advantages of fuzzy sets [7] and rough sets [1]. Fuzzy rough sets can handle continuous data without discretization. Fuzzification can effectively lower the information loss of data caused by discretization [8]–[10]. Subsequently, fuzzy covering rough sets, as a bridge between covering-based rough sets and fuzzy rough sets, have attracted the attention of many researchers [11]–[14]. Deng et al. [11] defined a novel fuzzy rough sets with a fuzzy covering based on the lattice theory. Li et al. [12] constructed two pairs of generalized fuzzy rough approximation operators via the concept of fuzzy covering. Feng et al. [13] investigated the reduction and fusion problem of fuzzy covering systems. D'eer et al. [14] extended 16 different fuzzy neighborhood operators based on fuzzy coverings and studied the relationship among them. However, the application condition of fuzzy covering is too harsh, which severely limits its development. Therefore, Ma [15] introduced the concept of fuzzy β-covering to solve this problem. After that, some researchers [16]–[20] studied fuzzy β-coverings. In theory, Huang et al. [16] established an intuitionistic fuzzy graded covering rough set model, which is a generalization of fuzzy β-covering rough sets [15] and intuitionistic fuzzy rough sets [21]. Subsequently, Yang and Hu [17] constructed four fuzzy β-neighborhood operators with fuzzy β-covering. Afterward, Zhang and Wang [18] further studied some problems existing in fuzzy β-covering approximation spaces (FβCAS). In terms of application, Jiang et al. [19] presented some of variable precision (I, T)-fuzzy rough set models with fuzzy β-covering for multiattribute decision making. Based on fuzzy covering rough sets with fuzzy β-covering, Zhang et al. [20] developed a novel multiattribute decision-making method by utilizing fuzzy measure and Choquet integral. In machine learning, the performance of classification task depends on the model and data. Data determine the upper bound of model performance, and model only approximates this bound. Feature engineering is a very important work for machine learning tasks. As a key step in feature engineering, the research of attribute reduction [22], [23] has been widely concerned. In general terms, attribute reduction is a method of data preprocessing. It can effectively improve the classification performance of a model.

Recently, the application of approximate reduction theory based on fuzzy β-covering has attracted extensive attention

[24]–[27]. Yang et al. [24] proposed a new granular reduction method with fuzzy β-coverings called the granular matrix. Subsequently, Huang and Li [25] first constructed a fitting model with fuzzy β-covering for attribute reduction. Then, Huang and Li [26] also defined a new discernibility measure and designed a fuzzy β-covering reduction algorithm for removing redundant fuzzy coverings. In addition, Huang et al. [27] presented a new multigranulation rough set model based on the noise-tolerant fuzzy βcovering and use it for feature subset selection. However, these existing methods with fuzzy β-covering for approximate reduction have some defects. For example, a general approach to obtain fuzzy β-coverings is lacking, the properties of fuzzy β-neighborhood operators for portraying similarity between samples are not sufficiently well developed, the conditional discrimination measure with fuzzy β-covering does not satisfy the monotonicity, and so on. Thus, we propose the notion of multifuzzy β-covering approximation spaces (MFβCAS) and its related theories for approximate reduction. The main motivations for conducting this research are shown as follows.

1) Since Ma [15] proposed the theory of fuzzy β-covering, a problem has been around. In detail, the definition of fuzzy β-covering is too strict, which makes it difficult to find fuzzy β-covering directly from the real data [24]–[27]. This means that fuzzy β-covering theory is difficult to be directly used to solve practical problems in life. The application of fuzzy β-covering theory is greatly restricted. Thus, the theory of fuzzy β-covering needs to be further refined and perfected.
2) From literature works [25] and [27], we can see that the authors used the original fuzzy β-neighborhood operator proposed by Ma [15] to characterize the similarity between samples. However, the original fuzzy β-neighborhood operator does not satisfy some properties necessary to characterize the relation between objects, including reflexivity, symmetry, and so on. Thus, it is unreasonable to depict the similarity between samples. Fortunately, Zhang et al. [28] proposed a reflexive fuzzy β-neighborhood operator based on R-implicator, which provides a new idea for constructing fuzzy β-neighborhood operators that satisfy the necessary conditions for characterizing the similarity between samples. For this reason, some new fuzzy β-neighborhood operators need to be constructed or a new fuzzy similarity relation based on fuzzy β-covering needs to be employed for describing the similarity between samples.
3) As mentioned above, the existing conditional discrimination index with fuzzy β-covering [26] has a theoretical defect, i.e., it does not satisfy monotonicity. This may lead to the instability of fuzzy β-covering reduction algorithm based on the conditional discrimination index. So, we need to put forward a new monotone conditional measure for designing a more stable fuzzy β-covering reduction algorithm.

To overcome these shortcomings, we first propose the notion of MFβCAS and explore its theoretical properties. In this way, the first problem that fuzzy β-covering is difficult to obtain from the real data is addressed by fuzzy information granulation. Besides, we introduce the fuzzy β-covering relation with reflexivity and symmetry to characterize the similarity between samples in the context of MFβCAS. Then, the second problem is solved. Finally, we present a new conditional discernibility measure with monotonicity for fuzzy β-covering reduction in the context of multifuzzy β-covering decision tables (MFβCDT). So far, the three defects have been made up.

The rest of this article is structured as follows. In Section II, we briefly review some basic knowledge that is relevant to this article. In Section III, we propose the notion of MFβCAS and discuss some of its properties. Section IV defines several information measures in the context of MFβCAS. Section V introduces a novel uncertainty measure for fuzzy β-covering reduction and designs a new framework of attribute reduction. Besides, we conduct a series of experiments to verify the effectiveness and stability of our proposed method in Section VI. Finally, Section VII summarizes this article.

II. Basic Notions

In this section, some basic concepts related to fuzzy β-covering and the knowledge related to information entropy are introduced. In addition, inspired by fuzzy similarity relation, the notion of fuzzy β-covering relation is also defined.

A. Fuzzy β-Covering Relations

First of all, let us review the related concepts of fuzzy β-covering and the definitions of several fuzzy β-neighborhood operators.

Definition 1 See ([15], [17], and [29]) (Fuzzy β-covering): Let $U = \{x_1, x_2, \ldots, x_i, \ldots, x_m\}$ be a nonempty universe, and $\mathcal{F}(U)$ be the family of all fuzzy sets on U. For any $\beta \in (0, 1]$ and $x \in U$, if $(\bigcup_{j=1}^{n} C_j)(x) \geqslant \beta$, then the family of fuzzy sets $\mathcal{C} = \{C_1, C_2, \ldots, C_j, \ldots, C_n\}$ is defined as a fuzzy β-covering of U, where $C_j \in \mathcal{F}(U)$. In addition, $< U, \mathcal{C} >$ is called an FβCAS. For any $x \in U$, the fuzzy β-neighborhood system $NS_{\mathcal{C}}^{\beta}(x)$ and fuzzy β-minimal description $md_{\mathcal{C}}^{\beta}(x)$ of x are defined as

$$NS_{\mathcal{C}}^{\beta}(x) = \{C \in \mathcal{C} \mid C(x) \geqslant \beta\}$$

$$md_{\mathcal{C}}^{\beta}(x) = \{C \in NS_{\mathcal{C}}^{\beta}(x) \mid \forall D \in NS_{\mathcal{C}}^{\beta}(x), D \subseteq C \Rightarrow C = D\}.$$

Proposition 1 (See [17]): Let $< U, \mathcal{C} >$ be an FβCAS. For any $x \in U$, $NS_{\mathcal{C}}^{\beta}(x) \supseteq md_{\mathcal{C}}^{\beta}(x)$ and $\bigcap NS_{\mathcal{C}}^{\beta}(x) = \bigcap md_{\mathcal{C}}^{\beta}(x)$.

Before presenting the existing fuzzy β-neighborhood operators, the concept of fuzzy β-neighborhood operators is given.

A fuzzy neighborhood operator [14] is a mapping N: $U \to \mathcal{F}(U)$. In particular, in the context of FβCASs, fuzzy neighborhood operators are called fuzzy β-neighborhood operators.

Based on the concept of fuzzy β-neighborhood operators, the fuzzy β-neighborhood operator $\tilde{N}_{\mathcal{C}}^{\beta}$ related to the fuzzy β-covering $\mathcal{C}$ is proposed by Ma [15], which is as follows.

Definition 2 (See [15]): Let $< U, \mathcal{C} >$ be an FβCAS. For any $x \in U$, the fuzzy β-neighborhood $\tilde{N}_{\mathcal{C}}^{\beta}(x)$ of x is defined as

$$\tilde{N}_{\mathcal{C}}^{\beta}(x) = \bigcap\{C \in \mathcal{C} \mid C(x) \geqslant \beta\}.$$

Subsequently, Zhang et al. [28] followed Ma's work and proposed a reflexive fuzzy β-neighborhood operator $\bar{N}_{\mathcal{C}}^{\beta}$ related to the fuzzy β-covering $\mathcal{C}$, which is given as follows.

Definition 3 (See [28]): Let $< U, \mathcal{C} >$ be an FβCAS. For any $x, y \in U$, the fuzzy β-neighborhood $\bar{N}_{\mathcal{C}}^{\beta}(x)$ of x is defined as

$$\bar{N}_{\mathcal{C}}^{\beta}(x)(y) = \underset{C \in md_{\mathcal{C}}^{\beta}(x)}{\wedge} \mathcal{I}(C(x), C(y))$$

where $\mathcal{I}$ is an R-implicator.

In machine learning, measuring the similarity between samples is a key step for many classification methods. Fuzzy similarity relations are usually employed to measure the similarity between samples in granular computing [6], [8], [9].

Definition 4 (Fuzzy similarity relation [6], [9]): Let U be a nonempty universe, $\mathcal{A}$ be an attribute set on U, and $\mathcal{B} \subseteq \mathcal{A}$. Suppose that $R_\mathcal{B}$ is a fuzzy binary relation deduced by attribute subset $\mathcal{B}$. For any $x, y \in U$, if $R_\mathcal{B}$ satisfies the conditions

1) reflexivity (i.e., $R_\mathcal{B}(x,x) = 1$)
2) symmetry (i.e., $R_\mathcal{B}(x,y) = R_\mathcal{B}(y,x)$)

then $R_\mathcal{B}$ is called a fuzzy similarity relation.

By the definition of fuzzy β-neighborhood operators, we can determine that a fuzzy β-neighborhood operator is a special fuzzy binary relation. Inspired by fuzzy similarity relation, the notion of fuzzy β-covering relation is also proposed.

Definition 5 (Fuzzy β-covering relation): Let $< U, \mathcal{C} >$ be an FβCAS, and $\mathcal{B} \subseteq \mathcal{C}$. Suppose that $R_\mathcal{B}^\beta$ is a fuzzy β-neighborhood operator related to the fuzzy β-covering $\mathcal{B}$. For any $x, y \in U$, if $R_\mathcal{B}^\beta$ satisfies the conditions

1) reflexivity (i.e., $R_\mathcal{B}^\beta(x,x) = 1$)
2) symmetry (i.e., $R_\mathcal{B}^\beta(x,y) = R_\mathcal{B}^\beta(y,x)$)

then $R_\mathcal{B}^\beta$ is called a fuzzy β-covering relation.

It is worth noting that fuzzy β-covering relation is a collective term for all fuzzy β-neighborhood operators with reflexivity and symmetry. In the context of FβCASs, the fuzzy similarity class $[x]_\mathcal{B}^\beta$ of x related to $\mathcal{B}$ is a fuzzy set on U for describing the similarity degrees between the sample x and all samples on domain U under knowledge $\mathcal{B}$. If $R_\mathcal{B}^\beta$ is a fuzzy β-covering relation, then $[x]_\mathcal{B}^\beta$ is called the fuzzy β-neighborhood of x related to $\mathcal{B}$, i.e., $[x]_\mathcal{B}^\beta(y) = R_\mathcal{B}^\beta(x,y)$ for any $y \in U$.

On this basis, several simple fuzzy β-neighborhood operators with reflexivity and symmetry are presented.

Definition 6: Let $< U, \mathcal{C} >$ be an FβCAS, $\mathcal{T}$ be a t-norm, and $\mathcal{S}$ be a t-conorm. For any $x, y \in U$, the four kinds of fuzzy β-neighborhoods ${}^1\mathrm{N}_\mathcal{C}^\beta(x)$, ${}^2\mathrm{N}_\mathcal{C}^\beta(x)$, ${}^3\mathrm{N}_\mathcal{C}^\beta(x)$, and ${}^4\mathrm{N}_\mathcal{C}^\beta(x)$ of x are, respectively, defined as

$$
\begin{aligned}
{}^1\mathrm{N}_\mathcal{C}^\beta(x)(y) &= \mathcal{T}(\bar{N}_\mathcal{C}^\beta(x)(y), \bar{N}_\mathcal{C}^\beta(y)(x)) \\
{}^2\mathrm{N}_\mathcal{C}^\beta(x)(y) &= \mathcal{S}(\bar{N}_\mathcal{C}^\beta(x)(y), \bar{N}_\mathcal{C}^\beta(y)(x)) \\
{}^3\mathrm{N}_\mathcal{C}^\beta(x)(y) &= {}^1\mathrm{N}_\mathcal{C}^\beta(x)(y) \wedge {}^2\mathrm{N}_\mathcal{C}^\beta(x)(y) \\
{}^4\mathrm{N}_\mathcal{C}^\beta(x)(y) &= {}^1\mathrm{N}_\mathcal{C}^\beta(x)(y) \vee {}^2\mathrm{N}_\mathcal{C}^\beta(x)(y).
\end{aligned}
$$

According to Definitions 5 and 6, we can determine that the four fuzzy β-neighborhood operators with reflexivity and symmetry ${}^1\mathrm{N}_\mathcal{B}^\beta$, ${}^2\mathrm{N}_\mathcal{B}^\beta$, ${}^3\mathrm{N}_\mathcal{B}^\beta$, and ${}^4\mathrm{N}_\mathcal{B}^\beta$ are fuzzy β-covering relations for $\mathcal{B} \subseteq \mathcal{C}$. For example, ${}^1\mathrm{N}_\mathcal{B}^\beta$ is a fuzzy β-covering relation, i.e., $R_\mathcal{B}^\beta = {}^1\mathrm{N}_\mathcal{B}^\beta$.

In order to make it easier to understand the concepts defined above, an example is given as follows.

Example 1: Let $< U, \mathcal{C} >$ be an FβCAS, where $U = \{x_1, x_2, x_3, x_4, x_5\}$, $\mathcal{C} = \{C_1, C_2, C_3, C_4\}$ is a fuzzy β-covering of U, and

$$
\begin{aligned}
C_1 &= \frac{0.5}{x_1} + \frac{0.7}{x_2} + \frac{0.8}{x_3} + \frac{0.6}{x_4} + \frac{0.7}{x_5}, \\
C_2 &= \frac{0.6}{x_1} + \frac{0.7}{x_2} + \frac{0.8}{x_3} + \frac{0.7}{x_4} + \frac{0.9}{x_5} \\
C_3 &= \frac{0.8}{x_1} + \frac{0.6}{x_2} + \frac{0.7}{x_3} + \frac{0.8}{x_4} + \frac{0.7}{x_5}, \\
C_4 &= \frac{0.6}{x_1} + \frac{0.5}{x_2} + \frac{0.9}{x_3} + \frac{0.9}{x_4} + \frac{0.7}{x_5}.
\end{aligned}
$$

For $\mathcal{B} = \{C_1, C_4\} \subseteq \mathcal{C}$, let $\beta = 0.6$ and $R_\mathcal{B}^\beta = {}^1\mathrm{N}_\mathcal{B}^\beta$, where $\mathcal{T} = \mathcal{T}_M$[1] and $\mathcal{I} = \mathcal{I}_L$[2], then we have

$$
R_\mathcal{B}^\beta = \begin{bmatrix} 1.0 & 0.8 & 0.7 & 0.7 & 0.8 \\ 0.8 & 1.0 & 0.6 & 0.6 & 0.8 \\ 0.7 & 0.6 & 1.0 & 0.8 & 0.8 \\ 0.7 & 0.6 & 0.8 & 1.0 & 0.8 \\ 0.8 & 0.8 & 0.8 & 0.8 & 1.0 \end{bmatrix}.
$$

Thus, we obtain the fuzzy similarity class $[x_i]_\mathcal{B}^\beta$ of x_i $(i = 1, 2, 3, 4, 5)$ related to the fuzzy β-covering $\mathcal{B}$, which is shown as

$$
\begin{aligned}
[x_1]_\mathcal{B}^\beta &= \frac{1.0}{x_1} + \frac{0.8}{x_2} + \frac{0.7}{x_3} + \frac{0.7}{x_4} + \frac{0.8}{x_5}, \\
[x_2]_\mathcal{B}^\beta &= \frac{0.8}{x_1} + \frac{1.0}{x_2} + \frac{0.6}{x_3} + \frac{0.6}{x_4} + \frac{0.8}{x_5} \\
[x_3]_\mathcal{B}^\beta &= \frac{0.7}{x_1} + \frac{0.6}{x_2} + \frac{1.0}{x_3} + \frac{0.8}{x_4} + \frac{0.8}{x_5}, \\
[x_4]_\mathcal{B}^\beta &= \frac{0.7}{x_1} + \frac{0.6}{x_2} + \frac{0.8}{x_3} + \frac{1.0}{x_4} + \frac{0.8}{x_5} \\
[x_5]_\mathcal{B}^\beta &= \frac{0.8}{x_1} + \frac{0.8}{x_2} + \frac{0.8}{x_3} + \frac{0.8}{x_4} + \frac{1.0}{x_5}.
\end{aligned}
$$

Essentially, fuzzy β-covering relation is a special fuzzy binary relation, which can be used to characterize the similarity between samples. Therefore, we will discuss some new uncertainty measures based on fuzzy β-covering relation in this article.

B. Information Entropy

Let U be a universe, and A be the set of all attributes in a dataset. Suppose that E and F are two attribute subsets of A. If $U/E = \{X_1, X_2, \ldots, X_i, \ldots, X_m\}$ and $U/F = \{Y_1, Y_2, \ldots, Y_j, \ldots, Y_n\}$ are two partitions derived by two equivalence relations R_E and R_F, then the probability distributions of E and F are, respectively, shown as follows:

$$
E \sim \begin{pmatrix} X_1 & X_2 & \ldots & X_i & \ldots & X_m \\ p(X_1) & p(X_2) & \ldots & p(X_i) & \ldots & p(X_m) \end{pmatrix} \tag{1}
$$

$$
F \sim \begin{pmatrix} Y_1 & Y_2 & \ldots & Y_j & \ldots & Y_n \\ p(Y_1) & p(Y_2) & \ldots & p(Y_j) & \ldots & p(Y_n) \end{pmatrix} \tag{2}
$$

where $p(X_i) = |X_i|/|U|$, $p(Y_j) = |Y_j|/|U|$ and $|\cdot|$ denotes the cardinality of a set.

Suppose that U is a universe, and $U/E = \{X_1, X_2, \ldots, X_i, \ldots, X_m\}$ is a partition of U related to E. If E has the probability distribution (1), then the information entropy of attribute subset E is defined as

$$
H(E) = -\sum_{i=1}^{m} p(X_i) \log p(X_i). \tag{3}
$$

[1]The standard min operator $\mathcal{T}_M(x,y) = \min\{x,y\}$ for any $x, y \in [0,1]$ is a common t-norm.

[2]The Łukasiewicz implicator $\mathcal{I}_L(x,y) = \min\{1, 1-x+y\}$ for any $x, y \in [0,1]$ is a common R-implicator.

Suppose that U is a universe, $U/E = \{X_1, X_2, \ldots, X_i, \ldots, X_m\}$ and $U/F = \{Y_1, Y_2, \ldots, Y_j, \ldots, Y_n\}$ are two partitions of U related to E and F, respectively. If E and F have the probability distributions (1) and (2), respectively, then the joint entropy, conditional entropy, and mutual information of attribute subsets E and F are defined as

$$H(E \cup F) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p(X_i \cap Y_j) \log p(X_i \cap Y_j) \quad (4)$$

$$H(E \mid F) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p(X_i \cap Y_j) \log p(X_i \mid Y_j) \quad (5)$$

$$I(E; F) = \sum_{i=1}^{m}\sum_{j=1}^{n} p(X_i \cap Y_j) \log \frac{p(X_i \cap Y_j)}{p(X_i)p(Y_j)} \quad (6)$$

where $p(X_i \cap Y_j) = |X_i \cap Y_j|/|U|$ and $p(X_i \mid Y_j) = |X_i \cap Y_j|/|Y_j|$.

In general, joint entropy is a generalization of entropy to multidimensional probability distributions, which describes the uncertainty of a set of random variables. Conditional entropy measures the amount of information of one random variable given the known value of another random variable. Mutual information reflects the degree of dependence of two random variables, that is, the degree of correlation. In the following, we will combine the theory of information entropy and the fuzzy β-covering relation to measure the uncertainty of fuzzy β-covering family.

III. Multifuzzy β-Covering Approximation Spaces

All fuzzy β-covering rough set models are constructed under a corresponding FβCAS. Therefore, if we want to utilize a fuzzy β-covering rough set model to solve a practical problem, this real problem must be transformed into a problem in an FβCAS. However, fuzzy β-covering is difficult to find directly from the real data. Fortunately, fuzzy information granulation provides a reasonable and effective way to obtain fuzzy β-coverings from the real data. Since fuzzy information granulation is capable of generating multiple fuzzy β-coverings, we introduce the notion of MFβCAS. First, the definition of MFβCAS is given as follows.

Definition 7: Let U be a nonempty universe and $\mathcal{F}(U)$ be the family of all fuzzy sets on U. Suppose that $\mathcal{M} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_j, \ldots, \mathcal{C}_n\}$ is a family of fuzzy β-coverings of U, where $\mathcal{C}_j = \{C_1^j, C_2^j, \ldots, C_i^j, \ldots, C_m^j\}$ is a fuzzy β-covering of U and $C_i^j \in \mathcal{F}(U)$, then the tuple $< U, \mathcal{M} >$ is called an MFβCAS).

Remark 1: In the context of multiattribute group decision making, we call the MFβCASs the fuzzy β-covering group approximation spaces in [30]. By Definitions 1 and 7, we can determine that the MFβCAS degenerates to an FβCAS when there is and only one fuzzy β-covering in an MFβCAS, that is, FβCASs are a special case of MFβCASs.

In order to make the fuzzy β-covering theory capable of solving some practical problems, the notion of MFβCASs has been presented above. In addition, some related definitions and properties of MFβCASs are given here. Then, the definition of fuzzy β-neighborhood in an MFβCAS is given in the following.

Definition 8: Let $< U, \mathcal{M} >$ be an MFβCAS and $\mathcal{G} \subseteq \mathcal{M}$. For any $x, y \in U$, the fuzzy β-neighborhood $[x]_{\mathcal{G}}^{\beta}$ of x related to $\mathcal{G}$ is defined as

$$[x]_{\mathcal{G}}^{\beta}(y) = \underset{\mathcal{C}\in\mathcal{G}}{\wedge} R_{\mathcal{C}}^{\beta}(x, y). \quad (7)$$

Note that $\mathrm{Cov}(\mathcal{G}) = \{[x]_{\mathcal{G}}^{\beta} \mid x \in U\}$ is a fuzzy β-covering of U induced by $\mathcal{G}$.

Considering that data noise may cause weak relationships between samples, a parameterized fuzzy β-neighborhood is given as

$$[x]_{\mathcal{G}}^{\beta,\lambda}(y) = \begin{cases} 0, & [x]_{\mathcal{G}}^{\beta}(y) < \lambda \\ [x]_{\mathcal{G}}^{\beta}(y), & [x]_{\mathcal{G}}^{\beta}(y) \geqslant \lambda \end{cases} \quad (8)$$

where λ is the radius of fuzzy β-neighborhood of a sample and $\lambda \in [0, 1]$.

From (8), we can see that there are three factors impacting the membership degrees of samples to the parameterized fuzzy β-neighborhood. The first is the fuzzy β-covering family $\mathcal{G}$, the second is the covering threshold β, and the third is the neighborhood radius λ.

Assume that $< U, \mathcal{M} >$ is an MFβCAS, $0 < \beta \leqslant 1$ and $0 \leqslant \lambda \leqslant 1$, $\mathcal{G} \subseteq \mathcal{M}$, and $[x]_{\mathcal{G}}^{\beta,\lambda}$ is the fuzzy β-neighborhood of x related to $\mathcal{G}$ and λ for any $x \in U$. Based on the above definitions, we can easily obtain the following properties.

Proposition 2: Let $\beta_1 \leqslant \beta_2$, then $[x]_{\mathcal{G}}^{\beta_1,\lambda} \subseteq [x]_{\mathcal{G}}^{\beta_2,\lambda}$.

Proof: For any $\mathcal{C} \in \mathcal{G}$, by Definitions 1 and 5, we can determine that the value of $R_{\mathcal{C}}^{\beta}$ becomes larger as the value of βincreases, i.e., $R_{\mathcal{C}}^{\beta_1} \subseteq R_{\mathcal{C}}^{\beta_2}$ when $\beta_1 \leqslant \beta_2$. Further, based on Definition 8 and (8), we have $[x]_{\mathcal{G}}^{\beta_1,\lambda}(y) \leqslant [x]_{\mathcal{G}}^{\beta_2,\lambda}(y)$ for any $y \in U$, that is, $[x]_{\mathcal{G}}^{\beta_1,\lambda} \subseteq [x]_{\mathcal{G}}^{\beta_2,\lambda}$. ∎

Proposition 3: Let $\lambda_1 \leqslant \lambda_2$, then $[x]_{\mathcal{G}}^{\beta,\lambda_1} \supseteq [x]_{\mathcal{G}}^{\beta,\lambda_2}$.

Proof: For any $y \in U$, for one thing, $[x]_{\mathcal{G}}^{\beta,\lambda_2}(y) = 0$ when $[x]_{\mathcal{G}}^{\beta}(y) < \lambda_2$, which implies $[x]_{\mathcal{G}}^{\beta,\lambda_1}(y) \geqslant [x]_{\mathcal{G}}^{\beta,\lambda_2}(y)$. For another, $[x]_{\mathcal{G}}^{\beta,\lambda_1}(y) = [x]_{\mathcal{G}}^{\beta,\lambda_2}(y) = [x]_{\mathcal{G}}^{\beta}(y)$ when $[x]_{\mathcal{G}}^{\beta}(y) \geqslant \lambda_2$. Thus, $[x]_{\mathcal{G}}^{\beta,\lambda_1}(y) \geqslant [x]_{\mathcal{G}}^{\beta,\lambda_2}(y)$, that is, $[x]_{\mathcal{G}}^{\beta,\lambda_1} \supseteq [x]_{\mathcal{G}}^{\beta,\lambda_2}$. ∎

Proposition 4: Let $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{M}$, then $[x]_{\mathcal{G}_1}^{\beta,\lambda} \supseteq [x]_{\mathcal{G}_2}^{\beta,\lambda}$.

Proof: Based on Definition 8, we have $[x]_{\mathcal{G}_1}^{\beta}(y) = \wedge_{\mathcal{C}\in\mathcal{G}_1} R_{\mathcal{C}}^{\beta}(x, y)$ and $[x]_{\mathcal{G}_2}^{\beta}(y) = \wedge_{\mathcal{C}\in\mathcal{G}_2} R_{\mathcal{C}}^{\beta}(x, y)$ for any $y \in U$. Let $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $\mathcal{C} \in \mathcal{G}_2$ for any $\mathcal{C} \in \mathcal{G}_1$. Thus, $[x]_{\mathcal{G}_1}^{\beta}(y) \geqslant [x]_{\mathcal{G}_2}^{\beta}(y)$. According to the formula (8), we can obtain $[x]_{\mathcal{G}_1}^{\beta,\lambda}(y) \geqslant [x]_{\mathcal{G}_2}^{\beta,\lambda}(y)$, that is, $[x]_{\mathcal{G}_1}^{\beta,\lambda} \supseteq [x]_{\mathcal{G}_2}^{\beta,\lambda}$. ∎

The fuzzy β-neighborhood represents the distinguishing ability of fuzzy β-covering family. The finer the fuzzy β-neighborhood is, the greater the distinguishing ability of fuzzy β-covering family is. Further, based on the definition of fuzzy β-neighborhood, a novel fuzzy β-covering rough set model in an MFβCAS is introduced as follows.

Definition 9: Let $< U, \mathcal{M} >$ be an MFβCAS. $\mathcal{G}$ is a subset of $\mathcal{M}$, and $[x]_{\mathcal{G}}^{\beta,\lambda}$ is the fuzzy β-neighborhood of x related to $\mathcal{G}$ and λ for any $x \in U$. For any $X \in \mathcal{F}(U)$, the lower approximation $\underline{R}_{\mathcal{G}}^{\beta,\lambda}(X)$ of X and the upper approximation $\overline{R}_{\mathcal{G}}^{\beta,\lambda}(X)$ of X are, respectively, defined as

$$\underline{R}_{\mathcal{G}}^{\beta,\lambda}(X)(x) = \underset{y\in U}{\wedge}\{(1 - [x]_{\mathcal{G}}^{\beta,\lambda}(y)) \vee X(y)\} \quad \forall x \in U \quad (9)$$

$$\overline{R}_{\mathcal{G}}^{\beta,\lambda}(X)(x) = \bigvee_{y\in U}\{[x]_{\mathcal{G}}^{\beta,\lambda}(y) \wedge X(y)\} \quad \forall x \in U. \tag{10}$$

In particular, when $\underline{R}_{\mathcal{G}}^{\beta,\lambda}(X) = \overline{R}_{\mathcal{G}}^{\beta,\lambda}(X)$, X is a definable set; otherwise, X is a fuzzy rough set, and $(\underline{R}_{\mathcal{G}}^{\beta,\lambda}(X), \overline{R}_{\mathcal{G}}^{\beta,\lambda}(X))$ is called a fuzzy β-covering rough set model of X in an MFβCAS.

Remark 2: Definition 9 is a rational extension of some existing rough set models. Therefore, other models can be derived from this definition, which is shown as follows.

1) When the fuzzy β-neighborhood $[x]_{\mathcal{G}}^{\beta,\lambda}$ of x is replaced by the crisp equivalence class $[x]_{\mathcal{G}}$ of x with covering relation and X is a crisp set on U

$$\underline{R}_{\mathcal{G}}(X) = \{x \in U \mid [x]_{\mathcal{G}} \subseteq X\}$$
$$\overline{R}_{\mathcal{G}}(X) = \{x \in U \mid [x]_{\mathcal{G}} \cap X \neq \emptyset\}$$

which is called the covering-based rough set model [3].

2) When the fuzzy β-neighborhood $[x]_{\mathcal{G}}^{\beta,\lambda}$ of x is replaced by the fuzzy similarity class $[x]_{\mathcal{G}}$ of x with fuzzy similarity relation and X is a fuzzy set on U

$$\underline{R}_{\mathcal{G}}(X)(x) = \bigwedge_{y\in U}\{(1 - R_{\mathcal{G}}(x,y)) \vee X(y)\}$$
$$\overline{R}_{\mathcal{G}}(X)(x) = \bigvee_{y\in U}\{R_{\mathcal{G}}(x,y) \wedge X(y)\}$$

which is called the fuzzy rough set model [6].

From Remark 2, we can conclude that the fuzzy β-covering rough set model in Definition 9 is a rational generalization of the covering-based rough set model [3] and the fuzzy rough set model [6].

Suppose that $< U, \mathcal{M} >$ is an MFβCAS, where U is a universe and $\mathcal{M}$ is a family of fuzzy β-coverings. Let $U/D = \{D_1, D_2, \ldots, D_r\}$ represent r crisp equivalence classes divided by the decision attribute D on U, then $< U, \mathcal{M}, D >$ is called MFβCDT.

Given an MFβCDT $< U, \mathcal{M}, D >$ and $\mathcal{G} \in \mathcal{M}$, the fuzzy positive region of the decision attribute D related to the fuzzy β-covering family $\mathcal{G}$ for any $x \in U$ is defined as

$$\mathrm{POS}_{\mathcal{G}}^{\beta,\lambda}(D)(x) = \bigcup_{i=1}^{r} \underline{R}_{\mathcal{G}}^{\beta,\lambda}(D_i)(x) \tag{11}$$

and the fuzzy dependence function of the decision attribute D related to the fuzzy β-covering family $\mathcal{G}$ is defined as

$$f_{\mathcal{G}}^{\beta,\lambda}(D) = \frac{\sum_{x\in U} \mathrm{POS}_{\mathcal{G}}^{\beta,\lambda}(D)(x)}{|U|}. \tag{12}$$

Besides, the fuzzy dependence function is also called the dependence degree, which reflects the classification ability of fuzzy β-covering family.

Definition 10: Let $< U, \mathcal{M}, D >$ be an MFβCDT, $0 < \beta \leqslant 1$, $0 \leqslant \lambda \leqslant 1$, and $\mathcal{G} \subseteq \mathcal{M}$. For any $\mathcal{C} \in \mathcal{G}$, the fuzzy β-covering $\mathcal{C}$ is regarded as redundant in the fuzzy β-covering family $\mathcal{G}$ if $f_{\mathcal{G}-\{\mathcal{C}\}}^{\beta,\lambda}(D) = f_{\mathcal{G}}^{\beta,\lambda}(D)$; otherwise, $\mathcal{C}$ is regarded as indispensable in $\mathcal{G}$. If all $\mathcal{C}$ in $\mathcal{G}$ are indispensable, then $\mathcal{G}$ is regarded as independent.

Definition 11: Let $< U, \mathcal{M}, D >$ be an MFβCDT, $0 < \beta \leqslant 1$, $0 \leqslant \lambda \leqslant 1$, and $\mathcal{G} \subseteq \mathcal{M}$. The fuzzy β-covering family $\mathcal{G}$ is a reduct of $\mathcal{M}$ iff

1) $f_{\mathcal{G}}^{\beta,\lambda}(D) = f_{\mathcal{M}}^{\beta,\lambda}(D)$
2) $\forall \mathcal{C} \in \mathcal{G}, f_{\mathcal{G}-\{\mathcal{C}\}}^{\beta,\lambda}(D) < f_{\mathcal{G}}^{\beta,\lambda}(D)$.

In order to verify the effectiveness and rationality of Definitions 10 and 11, an example is given as follows.

Example 2: Given an MFβCDT $< U, \mathcal{M}, D >$, where $U = \{x_1, x_2, x_3, x_4\}$, $\mathcal{M} = \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4\}$ is a family of fuzzy β-coverings of U, $\mathcal{C}_j = \{C_1^j, C_2^j, C_3^j, C_4^j\}$, $\mathcal{C}_j \in \mathcal{M}$, $U/D = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}$ is a partition of U on D, and

$$C_1^1 = \frac{0.9}{x_1} + \frac{0.69}{x_2} + \frac{0.26}{x_3} + \frac{0}{x_4}, C_2^1 = \frac{0.69}{x_1} + \frac{0.9}{x_2} + \frac{0.04}{x_3} + \frac{0}{x_4}$$
$$C_3^1 = \frac{0.26}{x_1} + \frac{0.04}{x_2} + \frac{0.9}{x_3} + \frac{0}{x_4}, C_4^1 = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0}{x_3} + \frac{0.9}{x_4}$$
$$C_1^2 = \frac{0.9}{x_1} + \frac{0}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}, C_2^2 = \frac{0}{x_1} + \frac{0.9}{x_2} + \frac{0}{x_3} + \frac{0.9}{x_4}$$
$$C_3^2 = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.9}{x_3} + \frac{0}{x_4}, C_4^2 = \frac{0}{x_1} + \frac{0.9}{x_2} + \frac{0}{x_3} + \frac{0.9}{x_4}$$
$$C_1^3 = \frac{0.9}{x_1} + \frac{0.08}{x_2} + \frac{0.29}{x_3} + \frac{0}{x_4}, C_2^3 = \frac{0.08}{x_1} + \frac{0.9}{x_2} + \frac{0}{x_3} + \frac{0.29}{x_4}$$
$$C_3^3 = \frac{0.29}{x_1} + \frac{0}{x_2} + \frac{0.9}{x_3} + \frac{0}{x_4}, C_4^3 = \frac{0}{x_1} + \frac{0.29}{x_2} + \frac{0}{x_3} + \frac{0.9}{x_4}$$
$$C_1^4 = \frac{0.9}{x_1} + \frac{0.52}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}, C_2^4 = \frac{0.52}{x_1} + \frac{0.9}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}$$
$$C_3^4 = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.9}{x_3} + \frac{0.52}{x_4}, C_4^4 = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.52}{x_3} + \frac{0.9}{x_4}.$$

Suppose that $R_{\mathcal{C}}^{\beta} = {}^{1}\mathrm{N}_{\mathcal{C}}^{\beta}$, where $\mathcal{T} = \mathcal{T}_M$ and $\mathcal{I} = \mathcal{I}_L$. Let $\beta = 0.5$ and $\lambda = 0.1$, then we can find a reduct red of $\mathcal{M}$ as follows.

Based on the (8), we can obtain that

$$[x_1]_{\mathcal{C}_1}^{\beta,\lambda} = \frac{1}{x_1} + \frac{0.79}{x_2} + \frac{0.36}{x_3} + \frac{0}{x_4},$$
$$[x_2]_{\mathcal{C}_1}^{\beta,\lambda} = \frac{0.79}{x_1} + \frac{1}{x_2} + \frac{0.14}{x_3} + \frac{0}{x_4}$$
$$[x_3]_{\mathcal{C}_1}^{\beta,\lambda} = \frac{0.36}{x_1} + \frac{0.14}{x_2} + \frac{1}{x_3} + \frac{0}{x_4}, [x_4]_{\mathcal{C}_1}^{\beta,\lambda} = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0}{x_3} + \frac{1}{x_4}$$
$$[x_1]_{\mathcal{C}_2}^{\beta,\lambda} = \frac{1}{x_1} + \frac{0}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}, [x_2]_{\mathcal{C}_2}^{\beta,\lambda} = \frac{0}{x_1} + \frac{1}{x_2} + \frac{0}{x_3} + \frac{1}{x_4}$$
$$[x_3]_{\mathcal{C}_2}^{\beta,\lambda} = \frac{0}{x_1} + \frac{0}{x_2} + \frac{1}{x_3} + \frac{0}{x_4}, [x_4]_{\mathcal{C}_2}^{\beta,\lambda} = \frac{0}{x_1} + \frac{1}{x_2} + \frac{0}{x_3} + \frac{1}{x_4}$$
$$[x_1]_{\mathcal{C}_3}^{\beta,\lambda} = \frac{1}{x_1} + \frac{0.18}{x_2} + \frac{0.39}{x_3} + \frac{0}{x_4},$$
$$[x_2]_{\mathcal{C}_3}^{\beta,\lambda} = \frac{0.18}{x_1} + \frac{1}{x_2} + \frac{0}{x_3} + \frac{0.39}{x_4}$$
$$[x_3]_{\mathcal{C}_3}^{\beta,\lambda} = \frac{0.39}{x_1} + \frac{0}{x_2} + \frac{1}{x_3} + \frac{0}{x_4}, [x_4]_{\mathcal{C}_3}^{\beta,\lambda} = \frac{0}{x_1} + \frac{0.39}{x_2} + \frac{0}{x_3} + \frac{1}{x_4}$$
$$[x_1]_{\mathcal{C}_4}^{\beta,\lambda} = \frac{1}{x_1} + \frac{0.62}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}, [x_2]_{\mathcal{C}_4}^{\beta,\lambda} = \frac{0.62}{x_1} + \frac{1}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}$$
$$[x_3]_{\mathcal{C}_4}^{\beta,\lambda} = \frac{0}{x_1} + \frac{0}{x_2} + \frac{1}{x_3} + \frac{0.62}{x_4}, [x_4]_{\mathcal{C}_4}^{\beta,\lambda} = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.62}{x_3} + \frac{1}{x_4}.$$

According to the formula (12), we can determine that $f_{\{\mathcal{C}_1\}}^{\beta,\lambda}(D) = 0.79$, $f_{\{\mathcal{C}_2\}}^{\beta,\lambda}(D) = 0.50$, $f_{\{\mathcal{C}_3\}}^{\beta,\lambda}(D) = 0.61$, and $f_{\{\mathcal{C}_4\}}^{\beta,\lambda}(D) = 0.69$. Consequently, we have red $= \{\mathcal{C}_1\}$. Since $\mathcal{G} = \{\mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4\} \neq \varnothing$ and $f_{\{\mathcal{C}_1\}}^{\beta,\lambda}(D) - 0 = 0.79 > 0$, we need to continue the next step. We calculate that $f_{\{\mathcal{C}_1,\mathcal{C}_2\}}^{\beta,\lambda}(D) =$

1, $f^{\beta,\lambda}_{\{\mathcal{C}_1,\mathcal{C}_3\}}(D)=0.82$, and $f^{\beta,\lambda}_{\{\mathcal{C}_1,\mathcal{C}_4\}}(D)=1$. Subsequently, we obtain a new reduct red $=\{\mathcal{C}_1,\mathcal{C}_2\}$. Since $G=\{\mathcal{C}_3,\mathcal{C}_4\}\neq\varnothing$ and $f^{\beta,\lambda}_{\{\mathcal{C}_1,\mathcal{C}_2\}}(D)-f^{\beta,\lambda}_{\{\mathcal{C}_1\}}(D)=0.21>0$, we need further calculations. Similarly, we have $f^{\beta,\lambda}_{\{\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_3\}}(D)=1$ and $f^{\beta,\lambda}_{\{\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_4\}}(D)=1$. Thus, we can obtain the final reduct red $=\{\mathcal{C}_1,\mathcal{C}_2\}$ with $f^{\beta,\lambda}_{\{\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_3\}}(D)-f^{\beta,\lambda}_{\{\mathcal{C}_1,\mathcal{C}_2\}}(D)=0$ and $f^{\beta,\lambda}_{\{\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_4\}}(D)-f^{\beta,\lambda}_{\{\mathcal{C}_1,\mathcal{C}_2\}}(D)=0$.

Based on the above definitions and conclusions, we can determine that MFβCASs are a special extension of FβCASs actually. Similar to FβCAS, the uncertainty of data can also be measured in an MFβCAS. In the following, we study several information measures in an MFβCAS.

IV. Information Measures in an MFβCAS

As an important method to measure the significance of attributes, information entropy has been widely used in attribute reduction [31], [32]. Similarly, information entropy theory can be used to measure the significance of fuzzy β-coverings [26]. In this section, by combining fuzzy β-covering theory and information theory, we extend the information measure in a fuzzy probability approximation space proposed by Hu et al. [31] and the neighborhood discrimination index in a fuzzy approximation space proposed by Wang et al. [32] to the context of MFβCASs, respectively. However, the conditional entropies in these two kinds of information measures are not monotonous, which will affect the selection of effective fuzzy β-coverings, resulting in the degradation of the classification performance of the model. Therefore, we propose a new measure of monotone conditional entropy in an MFβCDT and apply it to fuzzy β-covering reduction.

A. Information Measures Based on Shannon's Entropy

Based on the definition of parameterized fuzzy β-neighborhood and Shannon's entropy, the information measure proposed by Hu et al. [31] is extended to the context of MFβCASs in this part.

Definition 12: Let $<U,\mathcal{M}>$ be an MFβCAS, and $\mathcal{G}\subseteq\mathcal{M}$. $[x]^{\beta,\lambda}_{\mathcal{G}}$ is the fuzzy β-neighborhood of x related to $\mathcal{G}$ and λ for any $x\in U$, where λ is the radius of fuzzy β-neighborhood. The information measure of $\mathcal{G}$ in the MFβCAS is defined as

$$H^{\beta,\lambda}(\mathcal{G})=-\frac{1}{|U|}\sum_{x\in U}\log\frac{|[x]^{\beta,\lambda}_{\mathcal{G}}|}{|U|}. \tag{13}$$

Theorem 1: Let $<U,\mathcal{M}>$ be an MFβCAS and $\mathcal{G}\subseteq\mathcal{M}$, then the information measure of $\mathcal{G}$ satisfies the following properties.

(1) If $\beta_1\leqslant\beta_2$, then $H^{\beta_1,\lambda}(\mathcal{G})\geqslant H^{\beta_2,\lambda}(\mathcal{G})$.

(2) If $\lambda_1\leqslant\lambda_2$, then $H^{\beta,\lambda_1}(\mathcal{G})\leqslant H^{\beta,\lambda_2}(\mathcal{G})$.

(3) If $\mathcal{G}_1\subseteq\mathcal{G}_2\subseteq\mathcal{M}$, then $H^{\beta,\lambda}(\mathcal{G}_1)\leqslant H^{\beta,\lambda}(\mathcal{G}_2)$.

Note that the information measure in Definition 12 degenerates into Shannon's one when the fuzzy β-neighborhood $[x]^{\beta,\lambda}_{\mathcal{G}}$ of x is replaced by the crisp equivalence class $[x]_{\mathcal{G}}$ of x for any $x\in U$.

From Definition 12, it is easy to know that $H^{\beta,\lambda}(\mathcal{G})\geqslant 0$. Besides, the value of entropy increases monotonically with the increase of distinguishing ability of fuzzy β-covering family in an MFβCAS. It means that the finer the partition is, the larger the entropy is, and the more significant the fuzzy β-covering family is. Next, the definitions of joint entropy and conditional entropy are shown in the following.

Definition 13: Let $<U,\mathcal{M}>$ be an MFβCAS, and $\mathcal{E},\mathcal{F}\subseteq\mathcal{M}$. $[x]^{\beta,\lambda}_{\mathcal{E}}$ and $[x]^{\beta,\lambda}_{\mathcal{F}}$ are two fuzzy β-neighborhoods of x for any $x\in U$. The joint entropy of $\mathcal{E}$ and $\mathcal{F}$ in the MFβCAS is defined as

$$H^{\beta,\lambda}(\mathcal{E}\cup\mathcal{F})=-\frac{1}{|U|}\sum_{x\in U}\log\frac{|[x]^{\beta,\lambda}_{\mathcal{E}}\cap[x]^{\beta,\lambda}_{\mathcal{F}}|}{|U|}. \tag{14}$$

Definition 14: Let $<U,\mathcal{M}>$ be an MFβCAS and $\mathcal{E},\mathcal{F}\subseteq\mathcal{M}$. $[x]^{\beta,\lambda}_{\mathcal{E}}$ and $[x]^{\beta,\lambda}_{\mathcal{F}}$ are two fuzzy β-neighborhoods of x for $x\in U$. The conditional entropy of $\mathcal{F}$ conditioned to $\mathcal{E}$ in the MFβCAS is defined as

$$H^{\beta,\lambda}(\mathcal{F}\mid\mathcal{E})=-\frac{1}{|U|}\sum_{x\in U}\log\frac{|[x]^{\beta,\lambda}_{\mathcal{F}}\cap[x]^{\beta,\lambda}_{\mathcal{E}}|}{|[x]^{\beta,\lambda}_{\mathcal{E}}|}. \tag{15}$$

According to [31] and [33], we can determine that the conditional entropy in the formula (15) is not monotonous with the size of fuzzy β-covering family $\mathcal{E}$.

Theorem 2: Let $<U,\mathcal{M}>$ be an MFβCAS, and $\mathcal{E},\mathcal{F}\subseteq\mathcal{M}$. $[x]^{\beta,\lambda}_{\mathcal{E}}$ and $[x]^{\beta,\lambda}_{\mathcal{F}}$ are two fuzzy β-neighborhoods of x for any $x\in U$. We can obtain the following properties.

(1) $H^{\beta,\lambda}(\mathcal{E}\cup\mathcal{F})\geqslant\max\{H^{\beta,\lambda}(\mathcal{E}),H^{\beta,\lambda}(\mathcal{F})\}$.

(2) $H^{\beta,\lambda}(\mathcal{F}\mid\mathcal{E})=H^{\beta,\lambda}(\mathcal{E}\cup\mathcal{F})-H^{\beta,\lambda}(\mathcal{E})$.

(3) If $\mathcal{E}\subseteq\mathcal{F}$, then $H^{\beta,\lambda}(\mathcal{E}\cup\mathcal{F})=H^{\beta,\lambda}(\mathcal{F})$ and $H^{\beta,\lambda}(\mathcal{E}\mid\mathcal{F})=0$.

To sum up, Theorems 1 and 2 provide a theoretical support for the proposed information entropies to measure the distinguishing ability of fuzzy β-covering family.

B. Information Measures Based on Neighborhood Discrimination Index

Based on the definition of parameterized fuzzy β-neighborhood and discrimination index, the information measure proposed by Wang et al. [32] is also extended to the context of MFβCASs in this part.

Definition 15: Let $<U,\mathcal{M}>$ be an MFβCAS, and $\mathcal{G}\subseteq\mathcal{M}$. $[x]^{\beta,\lambda}_{\mathcal{G}}$ is the fuzzy β-neighborhood of x related to $\mathcal{G}$ and λ for any $x\in U$, where λ is the radius of fuzzy β-neighborhood. The neighborhood discrimination index of $\mathcal{G}$ in the MFβCAS is defined as

$$H^{\beta,\lambda}(\mathcal{G})=\log\frac{|U|^2}{\sum_{x\in U}|[x]^{\beta,\lambda}_{\mathcal{G}}|}. \tag{16}$$

Theorem 3: Let $<U,\mathcal{M}>$ be an MFβCAS and $\mathcal{G}\subseteq\mathcal{M}$, then the neighborhood discrimination index of $\mathcal{G}$ satisfies the following properties.

(1) If $\beta_1\leqslant\beta_2$, then $H^{\beta_1,\lambda}(\mathcal{G})\geqslant H^{\beta_2,\lambda}(\mathcal{G})$.

(2) If $\lambda_1\leqslant\lambda_2$, then $H^{\beta,\lambda_1}(\mathcal{G})\leqslant H^{\beta,\lambda_2}(\mathcal{G})$.

(3) If $\mathcal{G}_1\subseteq\mathcal{G}_2\subseteq\mathcal{M}$, then $H^{\beta,\lambda}(\mathcal{G}_1)\leqslant H^{\beta,\lambda}(\mathcal{G}_2)$.

From Definition 15, it is easy to know that $H^{\beta,\lambda}(\mathcal{G})\geqslant 0$. The neighborhood discrimination index in Definition 15 is used to measure the distinguishing ability of fuzzy β-covering family. Further, the joint discrimination index, conditional discrimination index, and mutual discrimination index are, respectively, introduced in the following.

Definition 16: Let $<U,\mathcal{M}>$ be an MFβCAS, and $\mathcal{E},\mathcal{F}\subseteq\mathcal{M}$. $[x]^{\beta,\lambda}_{\mathcal{E}}$ and $[x]^{\beta,\lambda}_{\mathcal{F}}$ are two fuzzy β-neighborhoods of x for

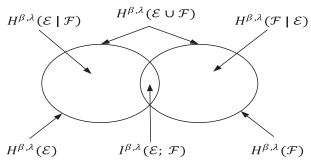


Fig. 1. Relationship between four discrimination indexes.

any $x \in U$. The joint discrimination index, conditional discrimination index, and mutual discrimination index of $\mathcal{E}$ and $\mathcal{F}$ in the MFβCAS are, respectively, defined as

$$H^{\beta,\lambda}(\mathcal{E} \cup \mathcal{F}) = \log \frac{|U|^2}{\sum_{x\in U} |[x]_{\mathcal{E}}^{\beta,\lambda} \cap [x]_{\mathcal{F}}^{\beta,\lambda}|} \tag{17}$$

$$H^{\beta,\lambda}(\mathcal{F} \mid \mathcal{E}) = \log \frac{\sum_{x\in U} |[x]_{\mathcal{E}}^{\beta,\lambda}|}{\sum_{x\in U} |[x]_{\mathcal{E}}^{\beta,\lambda} \cap [x]_{\mathcal{F}}^{\beta,\lambda}|} \tag{18}$$

$$I^{\beta,\lambda}(\mathcal{E}; \mathcal{F}) = \log \frac{|U|^2 \sum_{x\in U} |[x]_{\mathcal{E}}^{\beta,\lambda} \cap [x]_{\mathcal{F}}^{\beta,\lambda}|}{\sum_{x\in U} |[x]_{\mathcal{E}}^{\beta,\lambda}| \cdot \sum_{x\in U} |[x]_{\mathcal{F}}^{\beta,\lambda}|}. \tag{19}$$

According to [26] and [32], we can determine that the conditional discrimination index in (18) is also not monotonous with the size of fuzzy β-covering family $\mathcal{E}$.

Theorem 4: Let $< U, \mathcal{M} >$ be an MFβCAS, and $\mathcal{E}, \mathcal{F} \subseteq \mathcal{M}$. $[x]_{\mathcal{E}}^{\beta,\lambda}$ and $[x]_{\mathcal{F}}^{\beta,\lambda}$ are two fuzzy β-neighborhoods of x for any $x \in U$. We can obtain the following properties.

1) $H^{\beta,\lambda}(\mathcal{E} \cup \mathcal{F}) \geqslant \max\{H^{\beta,\lambda}(\mathcal{E}), H^{\beta,\lambda}(\mathcal{F})\}$.
2) $H^{\beta,\lambda}(\mathcal{F} \mid \mathcal{E}) = H^{\beta,\lambda}(\mathcal{E} \cup \mathcal{F}) - H^{\beta,\lambda}(\mathcal{E})$.
3) If $\mathcal{E} \subseteq \mathcal{F}$, then $H^{\beta,\lambda}(\mathcal{E} \cup \mathcal{F}) = H^{\beta,\lambda}(\mathcal{F})$ and $H^{\beta,\lambda}(\mathcal{E} \mid \mathcal{F}) = 0$.
4) $I^{\beta,\lambda}(\mathcal{E}; \mathcal{F}) = I^{\beta,\lambda}(\mathcal{F}; \mathcal{E})$.
5) $I^{\beta,\lambda}(\mathcal{E}; \mathcal{F}) = H^{\beta,\lambda}(\mathcal{E}) + H^{\beta,\lambda}(\mathcal{F}) - H^{\beta,\lambda}(\mathcal{E} \cup \mathcal{F})$.
6) $I^{\beta,\lambda}(\mathcal{E}; \mathcal{F}) = H^{\beta,\lambda}(\mathcal{E}) - H^{\beta,\lambda}(\mathcal{E} \mid \mathcal{F}) = H^{\beta,\lambda}(\mathcal{F}) - H^{\beta,\lambda}(\mathcal{F} \mid \mathcal{E})$.

In order to more intuitively understand the relationship between the neighborhood, joint, conditional, and mutual discrimination indexes in an MFβCAS, the graph of the relationship between four discrimination indexes is given in Fig. 1.

C. Novel Information Measure Based on the Monotone Conditional Entropy

From the previous content, we can see that the measures of conditional entropies in (15) and (18) do not satisfy monotonicity, which may cause the algorithms of fuzzy β-covering reduction to be unstable. Therefore, a novel measure of monotone conditional entropy in an MFβCDT is proposed as follows.

Definition 17: Let $< U, \mathcal{M}, D >$ be an MFβCDT, and $\mathcal{G} \subseteq \mathcal{M}$. $U/D = \{D_1, D_2, \ldots, D_l, \ldots, D_r\}$ is composed of r crisp equivalence classes divided by the decision attribute D on U. $[x]_{\mathcal{G}}^{\beta,\lambda}$ is the fuzzy β-neighborhood of x related to $\mathcal{G}$ and λ for any $x \in U$, where λ is the radius of fuzzy β-neighborhood. The monotone conditional entropy of D conditioned to $\mathcal{G}$ in the MFβCDT is defined as

$$H^{\beta,\lambda}(D \mid \mathcal{G}) = -\sum_{x\in U} \sum_{D_l \in U/D} \frac{|[x]_{\mathcal{G}}^{\beta,\lambda} \cap D_l|}{|U|} \log \frac{|[x]_{\mathcal{G}}^{\beta,\lambda} \cap D_l|}{|[x]_{\mathcal{G}}^{\beta,\lambda}|}. \tag{20}$$

Further, the probability form of monotone conditional entropy of D conditioned to $\mathcal{G}$ in an MFβCDT can be expressed as

$$H^{\beta,\lambda}(D \mid \mathcal{G}) = -\sum_{x\in U} p([x]_{\mathcal{G}}^{\beta,\lambda}) \sum_{D_l \in U/D} p(D_l \mid [x]_{\mathcal{G}}^{\beta,\lambda}) \log p(D_l \mid [x]_{\mathcal{G}}^{\beta,\lambda}) \tag{21}$$

where $p([x]_{\mathcal{G}}^{\beta,\lambda}) = |[x]_{\mathcal{G}}^{\beta,\lambda}|/|U|$ and $p(D_l \mid [x]_{\mathcal{G}}^{\beta,\lambda}) = |[x]_{\mathcal{G}}^{\beta,\lambda} \cap D_l|/|[x]_{\mathcal{G}}^{\beta,\lambda}|$.

By analyzing Definition 17, some important properties of the monotone conditional entropy in an MFβCDT are obtained, which are shown in the following.

Theorem 5 (Equivalence): Let $< U, \mathcal{M}, D >$ be an MFβCDT, and $\mathcal{E}, \mathcal{F} \subseteq \mathcal{M}$. $[x]_{\mathcal{E}}^{\beta,\lambda}$ and $[x]_{\mathcal{F}}^{\beta,\lambda}$ are two fuzzy β-neighborhoods of x for any $x \in U$. If $[x]_{\mathcal{E}}^{\beta,\lambda} = [x]_{\mathcal{F}}^{\beta,\lambda}$ for any $x \in U$, then $H^{\beta,\lambda}(D \mid \mathcal{E}) = H^{\beta,\lambda}(D \mid \mathcal{F})$.

Theorem 6 (Maximum): Let $< U, \mathcal{M}, D >$ be an MFβCDT, and $\mathcal{G} \subseteq \mathcal{M}$. The maximum of monotone conditional entropy of D conditioned to $\mathcal{G}$ is $|U| \log |U|$ only if the fuzzy β-neighborhood $|[x]_{\mathcal{G}}^{\beta,\lambda}| = |U|$ for any $x \in U$ and the decision equivalence class $|D_l| = 1$ for any $D_l \in U/D$, that is, $H_{\max}^{\beta,\lambda}(D \mid \mathcal{G}) = |U| \log |U|$.

Theorem 7 (Minimum): Let $< U, \mathcal{M}, D >$ be an MFβCDT, and $U/D = \{[x]_D \mid x \in U\}$ is a family of crisp equivalence classes on U divided by D. $[x]_{\mathcal{G}}^{\beta,\lambda}$ is the fuzzy β-neighborhood of x related to $\mathcal{G}$ and λ for any $x \in U, \mathcal{G} \subseteq \mathcal{M}$ and λ is the radius of fuzzy β-neighborhood. The minimum of monotone conditional entropy of D conditioned to $\mathcal{G}$ is 0 only if $[x]_{\mathcal{G}}^{\beta,\lambda} \subseteq [x]_D$ for any $x \in U$, that is, $H_{\min}^{\beta,\lambda}(D \mid \mathcal{G}) = 0$.

Proof: For any $x \in U$, $[x]_{\mathcal{G}}^{\beta,\lambda} \subseteq [x]_D$, we have

$$\begin{aligned} H^{\beta,\lambda}(D \mid \mathcal{G}) &= -\sum_{x\in U} \frac{|[x]_{\mathcal{G}}^{\beta,\lambda} \cap [x]_D|}{|U|} \log \frac{|[x]_{\mathcal{G}}^{\beta,\lambda} \cap [x]_D|}{|[x]_{\mathcal{G}}^{\beta,\lambda}|} \\ &= -\sum_{x\in U} \frac{|[x]_{\mathcal{G}}^{\beta,\lambda}|}{|U|} \log \frac{|[x]_{\mathcal{G}}^{\beta,\lambda}|}{|[x]_{\mathcal{G}}^{\beta,\lambda}|} = 0. \end{aligned}$$

■

Theorem 8 (Monotonicity): Let $< U, \mathcal{M}, D >$ be an MFβCDT and $\mathcal{G} \subseteq \mathcal{M}$, then the monotone conditional entropy of D satisfies the following properties.

1) If $\beta_1 \leqslant \beta_2$, then $H^{\beta_1,\lambda}(D \mid \mathcal{G}) \leqslant H^{\beta_2,\lambda}(D \mid \mathcal{G})$.
2) If $\lambda_1 \leqslant \lambda_2$, then $H^{\beta,\lambda_1}(D \mid \mathcal{G}) \geqslant H^{\beta,\lambda_2}(D \mid \mathcal{G})$.
3) If $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{M}$, then $H^{\beta,\lambda}(D \mid \mathcal{G}_1) \geqslant H^{\beta,\lambda}(D \mid \mathcal{G}_2)$.

Proof: It can easily proved by Propositions 2–4 and Definition 17. ■

Suppose that $< U, \mathcal{M}, D >$ is an MFβCDT, and $U/D = \{[x]_D \mid x \in U\}$ is a family of crisp equivalence classes on U divided by D. $[x]_{\mathcal{M}}^{\beta,\lambda}$ is the fuzzy β-neighborhood of x related to $\mathcal{M}$ and λ for any $x \in U$, and λ is the radius of fuzzy β-neighborhood. If $[x]_{\mathcal{M}}^{\beta,\lambda} \subseteq [x]_D$ for any $x \in U$, then $< U, \mathcal{M}, D >$ is a consistent MFβCDT.

Theorem 9: Let $< U, \mathcal{M}, D >$ be a consistent MFβCDT, then $H^{\beta,\lambda}(D \mid \mathcal{M}) = 0$.

Proof: It can easily proved by Theorem 7. ■

Based on the above theorems, we can ensure that the monotone conditional entropy in an MFβCDT proposed by us can be used as a reasonable uncertainty measure for fuzzy β-covering reduction. The smaller the monotone conditional entropy of fuzzy β-covering family is, the better the distinguishing ability of fuzzy β-covering family is, and the more significant the fuzzy β-covering family is.

In this article, some existing uncertain measures are extended to the context of MFβCASs by using the parameterized fuzzy β-neighborhood, which provides a variety of methods to measure the significance of fuzzy β-coverings. Moreover, in an MFβCDT, the proposed monotone conditional entropy can reasonably and effectively measure the significance of fuzzy β-coverings compared with other nonmonotone conditional entropies. In addition, the relationship among the newly proposed information measures and the information measures with fuzzy β-covering proposed by Huang and Li in [26] is further demonstrated. The connection between them is that these information measures are not only based on fuzzy β-neighborhood operators to characterize the similarity between samples but also based on information theory to measure the significance of fuzzy β-coverings. The difference between them is that, in [26], the original fuzzy β-neighborhood operator defined by Ma [15] is employed to characterize the similarity between samples, which does not meet the reflexivity and symmetry, while in this article, the fuzzy β-covering relation with reflexivity and symmetry is used to characterize the similarity between samples. Furthermore, the conditional discrimination measure in [26] does not meet the monotonicity, while the conditional information measure proposed in this article meets the monotonicity.

In order to further show how the concept given by Definition 17 in this article improves existing ones in [26], an example is given as follows.

Example 3 (Following Example 2): Given an MFβCDT $< U, \mathcal{M}, D >$, where $U = \{x_1, x_2, x_3, x_4\}$, $\mathcal{M} = \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4\}$ is a family of fuzzy β-coverings of U, $\mathcal{C}_j = \{C_1^j, C_2^j, C_3^j, C_4^j\}, \mathcal{C}_j \in \mathcal{M}$. $U/D = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}$ is a partition of U on D.

According to literature [26], let $\beta = 0.5$ and $\delta = 0.1$, then we have

$$[x_1]_{\mathcal{C}_1}^{\beta,\delta} = \frac{0.65}{x_1} + \frac{0.65}{x_2} + \frac{0}{x_3} + \frac{0}{x_4},$$

$$[x_2]_{\mathcal{C}_1}^{\beta,\delta} = \frac{0.65}{x_1} + \frac{0.65}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}$$

$$[x_3]_{\mathcal{C}_1}^{\beta,\delta} = \frac{0.16}{x_1} + \frac{0}{x_2} + \frac{0.9}{x_3} + \frac{0}{x_4}, [x_4]_{\mathcal{C}_1}^{\beta,\delta} = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0}{x_3} + \frac{0.9}{x_4}$$

$$[x_1]_{\mathcal{C}_2}^{\beta,\delta} = \frac{0.9}{x_1} + \frac{0}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}, [x_2]_{\mathcal{C}_2}^{\beta,\delta} = \frac{0}{x_1} + \frac{0.9}{x_2} + \frac{0}{x_3} + \frac{0.9}{x_4}$$

$$[x_3]_{\mathcal{C}_2}^{\beta,\delta} = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.9}{x_3} + \frac{0}{x_4}, [x_4]_{\mathcal{C}_2}^{\beta,\delta} = \frac{0}{x_1} + \frac{0.9}{x_2} + \frac{0}{x_3} + \frac{0.9}{x_4}$$

$$[x_1]_{\mathcal{C}_3}^{\beta,\delta} = \frac{0.9}{x_1} + \frac{0}{x_2} + \frac{0.19}{x_3} + \frac{0}{x_4}, [x_2]_{\mathcal{C}_3}^{\beta,\delta} = \frac{0}{x_1} + \frac{0.9}{x_2} + \frac{0}{x_3} + \frac{0.19}{x_4}$$

$$[x_3]_{\mathcal{C}_3}^{\beta,\delta} = \frac{0.19}{x_1} + \frac{0}{x_2} + \frac{0.9}{x_3} + \frac{0}{x_4}, [x_4]_{\mathcal{C}_3}^{\beta,\delta} = \frac{0}{x_1} + \frac{0.19}{x_2} + \frac{0}{x_3} + \frac{0.9}{x_4}$$

$$[x_1]_{\mathcal{C}_4}^{\beta,\delta} = \frac{0.9}{x_1} + \frac{0.46}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}, [x_2]_{\mathcal{C}_4}^{\beta,\delta} = \frac{0.46}{x_1} + \frac{0.9}{x_2} + \frac{0}{x_3} + \frac{0}{x_4}$$

$$[x_3]_{\mathcal{C}_4}^{\beta,\delta} = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.9}{x_3} + \frac{0.46}{x_4},$$

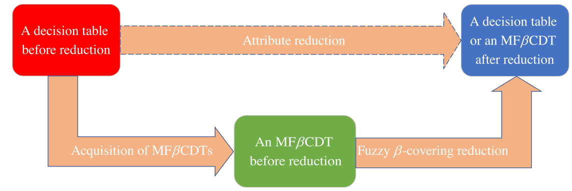


Fig. 2. Framework diagram of attribute reduction.

$$[x_4]_{\mathcal{C}_4}^{\beta,\delta} = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.46}{x_3} + \frac{0.9}{x_4}.$$

Further, we can obtain that $H^{\beta,\delta}(D \mid \mathcal{C}_1) = 0.0151$ and $H^{\beta,\delta}(D \mid \{\mathcal{C}_1, \mathcal{C}_3\}) = 0.0214$.

In this article, suppose that $R_{\mathcal{C}}^{\beta} = {}^1\mathrm{N}_{\mathcal{C}}^{\beta}$, where $\mathcal{T} = \mathcal{T}_M$ and $\mathcal{I} = \mathcal{I}_L$. Let $\beta = 0.5$ and $\lambda = 0.1$, then we can obtain $[x_i]_{\mathcal{C}_j}^{\beta,\lambda} (i = 1, 2, 3, 4; j = 1, 2, 3, 4)$ by Definition 17, which can be found from Example 2. Further, we can obtain that $H^{\beta,\lambda}(D \mid \mathcal{C}_1) = 0.2629$ and $H^{\beta,\lambda}(D \mid \{\mathcal{C}_1, \mathcal{C}_3\}) = 0.1751$.

In general, conditional entropy should decrease monotonically with the increase of attributes [34]. Similarly, in the context of MFβCDTs, conditional information measure should decrease monotonically with the increase of fuzzy β-coverings. According to Example 3, we can determine that our proposed conditional information measure satisfies this condition, while the existing one does not. Based on the above results, we can clearly see the difference between our proposed conditional information measure and the existing one.

V. Framework of Attribute Reduction Based on Fuzzy β-Covering Reduction

In machine learning, attribute reduction is an important data preprocessing process. Based on the importance measures of fuzzy β-covering proposed in the Section IV, we can design many different algorithms for fuzzy β-covering reduction. Moreover, an attribute is able to derive a fuzzy β-covering in a decision table via fuzzy information granulation. Therefore, we propose a framework of attribute reduction based on fuzzy β-covering reduction in this section. The framework diagram of attribute reduction is shown in Fig. 2.

From Fig. 2, we can see that the proposed framework has two steps for attribute reduction. The first step is the acquisition of MFβCDTs; the second step is the fuzzy β-covering reduction. In the following, we introduce each of these two steps in detail.

A. Acquisition of MFβCDTs via Fuzzy Information Granulation

A prerequisite for the proposed theory related to fuzzy β-covering to be applied to solve a practical problem is that a reasonable and valid MFβCDT can be obtained. Fortunately, fuzzy information granulation is able to extract effective information and discard redundant information from the raw data based on different levels of granularity and, thus, can solve the practical problem from different levels of granularity [35], [36]. It provides an effective and reasonable way to obtain MFβCDTs. In this article, we conduct fuzzy information granulation on the raw data by using fuzzy similarity relation so as to obtain the matrices of similarity between samples, that is, the effective classification information. Further, by introducing an adjustable

TABLE I
DECISION TABLE $< U, \mathcal{A}, D >$

$U/\mathcal{A}$	a_1	a_2	a_3	a_4	D
x_1	0.3	1	0.7	0.8	1
x_2	0.4	0.5	0.3	1	1
x_3	0	0	1	0	2
x_4	1	0.5	0	0.2	3

coefficient η of similarity and $\eta \in [0,1]$, we extend the fuzzy similarity relation to η-fuzzy similarity relation. Note that fuzzy similarity relation is a special case of η-fuzzy similarity relation. The definition of η-fuzzy similarity relation is given as follows:

$$R_B^{\eta}(x,y) = \eta * R_B(x,y) \quad \forall x,y \in U \tag{22}$$

where $R_B(x,y)$ is an arbitrary fuzzy similarity relation on domain U deduced by attribute subset B.

Based on the definitions of MFβCDTs and η-fuzzy similarity relations, we can determine that MFβCDTs can be obtained from the original decision tables through η-fuzzy similarity relations when $\eta \in [\beta, 1]$. Then, the detailed construction process of MFβCDTs is described as follows.

Let $< U, \mathcal{A}, D >$ be a decision table, where $U = \{x_1, x_2, \ldots, x_i, \ldots, x_m\}$ is a set of samples, $\mathcal{A} = \{a_1, a_2, \ldots, a_j, \ldots, a_n\}$ is a set of attributes, and D is a decision attribute. If $R_{\{a_j\}}$ is a fuzzy similarity relation on U induced by a_j, then the similarity between sample x and sample y related to attribute a_j is given as

$$R_{\{a_j\}}^{\eta}(x,y) = \eta * R_{\{a_j\}}(x,y) \quad \forall x,y \in U \tag{23}$$

where η is an adjustable coefficient and $\eta \in [\beta, 1]$. In real application, η is often set to 0.9 to make as many samples enter the positive domain as possible [26], [27]. For any $x_i, y \in U$, let $C_i^j(y) = R_{\{a_j\}}^{\eta}(x_i, y)$, then we have $C_i^j(x_i) = R_{\{a_j\}}^{\eta}(x_i, x_i) = \eta \geqslant \beta$ and $(\cup_{k=1}^{m} C_k^j)(x_i) \geqslant \beta$. Thus, $\mathcal{C}_j = \{C_1^j, C_2^j, \ldots, C_i^j, \ldots, C_m^j\}$ is a fuzzy β-covering induced by a_j. Consequently, an MFβCDT $< U, \mathcal{M}, D >$ is obtained, where $\mathcal{M} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_j, \ldots, \mathcal{C}_n\}$.

In order to demonstrate the construction process of MFβCDTs, an example is given as follows.

Example 4: Given a decision table $< U, \mathcal{A}, D >$ in Table I, where $U = \{x_1, x_2, x_3, x_4\}$, and $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$.

Let $\beta = 0.5$ and $\eta = 0.9$. Assume that the fuzzy similarity relation $R_{\{a_j\}}$ for single numerical attribute uses the following binary function [25], [37]:

$$R_{\{a_j\}}(x,y) = \max\left(\min\left(\frac{a_j(x) - a_j(y) + \sigma_{a_j}}{\sigma_{a_j}}, \frac{a_j(y) - a_j(x) + \sigma_{a_j}}{\sigma_{a_j}}\right), 0\right) \tag{24}$$

where σ_{a_j} is the standard deviation of attribute a_j.

From Table I, we can obtain that $\sigma_{a_1} = 0.42, \sigma_{a_2} = 0.41, \sigma_{a_3} = 0.44, \sigma_{a_4} = 0.48$. Based on the formula (23), we have

$$R_{\{a_1\}}^{0.9} = \begin{bmatrix} 0.9 & 0.69 & 0.26 & 0 \\ 0.69 & 0.9 & 0.04 & 0 \\ 0.26 & 0.04 & 0.9 & 0 \\ 0 & 0 & 0 & 0.9 \end{bmatrix}, \quad R_{\{a_2\}}^{0.9} = \begin{bmatrix} 0.9 & 0 & 0 & 0 \\ 0 & 0.9 & 0 & 0.9 \\ 0 & 0 & 0.9 & 0 \\ 0 & 0.9 & 0 & 0.9 \end{bmatrix}$$

$$R_{\{a_3\}}^{0.9} = \begin{bmatrix} 0.9 & 0.08 & 0.29 & 0 \\ 0.08 & 0.9 & 0 & 0.29 \\ 0.29 & 0 & 0.9 & 0 \\ 0 & 0.29 & 0 & 0.9 \end{bmatrix},$$

$$R_{\{a_4\}}^{0.9} = \begin{bmatrix} 0.9 & 0.52 & 0 & 0 \\ 0.52 & 0.9 & 0 & 0 \\ 0 & 0 & 0.9 & 0.52 \\ 0 & 0 & 0.52 & 0.9 \end{bmatrix}.$$

It follows that

$$C_1^1 = \frac{0.9}{x_1} + \frac{0.69}{x_2} + \frac{0.26}{x_3} + \frac{0}{x_4}, \quad C_2^1 = \frac{0.69}{x_1} + \frac{0.9}{x_2} + \frac{0.04}{x_3} + \frac{0}{x_4}$$

$$C_3^1 = \frac{0.26}{x_1} + \frac{0.04}{x_2} + \frac{0.9}{x_3} + \frac{0}{x_4}, \quad C_4^1 = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0}{x_3} + \frac{0.9}{x_4}.$$

It is obvious that $(\cup_{k=1}^{4} C_k^1)(x) \geqslant \beta$ for any $x \in U$. Thus, $\mathcal{C}_1 = \{C_1^1, C_2^1, C_3^1, C_4^1\}$ is a fuzzy β-covering induced by a_1. Similarly, $\mathcal{C}_2 = \{C_1^2, C_2^2, C_3^2, C_4^2\}$, $\mathcal{C}_3 = \{C_1^3, C_2^3, C_3^3, C_4^3\}$, and $\mathcal{C}_4 = \{C_1^4, C_2^4, C_3^4, C_4^4\}$ are three fuzzy β-coverings. Finally, an MFβCDT $< U, \mathcal{M}, D >$ is obtained, where $\mathcal{M} = \{\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4\}$.

B. Novel Uncertainty Measure for Fuzzy β-Covering Reduction

In this section, a heuristic fuzzy β-covering reduction algorithm based on the monotone conditional entropy in an MFβCDT is designed. When the monotone conditional entropy is regarded as a criterion for fuzzy β-covering reduction, the search stop step of algorithm can be easily implemented. In order to find the family of fuzzy β-coverings with maximum distinguishing ability and delete the redundant fuzzy β-coverings before classification learning, the definition of a reduct in an MFβCDT is given as follows.

Definition 18: Let $< U, \mathcal{M}, D >$ be an MFβCDT, $0 < \beta \leqslant 1$, $0 \leqslant \lambda \leqslant 1$, and $\mathcal{G} \subseteq \mathcal{M}$. The fuzzy β-covering family $\mathcal{G}$ is a reduct of $\mathcal{M}$ iff

1) $H^{\beta,\lambda}(D \mid \mathcal{G}) = H^{\beta,\lambda}(D \mid \mathcal{M})$
2) $\forall \mathcal{C} \in \mathcal{G}, H^{\beta,\lambda}(D \mid \mathcal{G} - \{\mathcal{C}\}) > H^{\beta,\lambda}(D \mid \mathcal{G})$.

From Definition 18, we can see that the minimal family of fuzzy β-coverings after reduction has the same classification ability with the original family of fuzzy β-coverings. Further, to evaluate the distinguishing ability of a new fuzzy β-covering family, the definition of significance of a fuzzy β-covering is given as follows.

Definition 19: Let $< U, \mathcal{M}, D >$ be an MFβCDT, $\mathcal{G} \subseteq \mathcal{M}$, and $\mathcal{C} \in \mathcal{M} - \mathcal{G}$, then the significance of $\mathcal{C}$ related to $\mathcal{G}$ for D is defined as

$$\mathrm{SIG}^{\beta,\lambda}(\mathcal{C}, \mathcal{G}, D) = H^{\beta,\lambda}(D \mid \mathcal{G}) - H^{\beta,\lambda}(D \mid \mathcal{G} \cup \{\mathcal{C}\}). \tag{25}$$

In particular, $H^{\beta,\lambda}(D \mid \mathcal{G}) = H_{\max}^{\beta,\lambda}(D \mid \mathcal{G})$ if $\mathcal{G} = \varnothing$. The significance of a new fuzzy β-covering depends on the increment of distinguishing information after adding the fuzzy β-covering into the selected family of fuzzy β-coverings. A large value of

Algorithm 1: Forward greedy fuzzy β-covering reduction algorithm based on the monotone conditional entropy (MFBC).

```
Input: An MFβCDT < U, M, D >, the covering threshold βand the neighborhood radius λ.
Output: One reduct red.
Initialize red = ∅, G = M;
for each C ∈ G do
  Compute the fuzzy β-covering relation R_C^β;
end for
while G ≠ ∅ do
  for each C ∈ G do
    for each x ∈ U do
      Compute the fuzzy β-neighborhood [x]_{red∪{C}}^{β,λ};
    end for
    Compute the monotone conditional entropy
    H^{β,λ}(D | red ∪ {C});
  end for
  Find C′ = C when H^{β,λ}(D | red ∪ {C}) is the
  minimum for each C ∈ G;
  Compute the significance SIG^{β,λ}(C,′ red, D) =
  H^{β,λ}(D | red) − H^{β,λ}(D | red ∪ {C′});
  if SIG^{β,λ}(C,′ red, D) > 0 then
    red ← red ∪ {C′};
    G ← G − red;
  else
    return red;
  end if
end while
return red
```

$\mathrm{SIG}^{\beta,\lambda}(\mathcal{C},\mathcal{G},D)$ implies that the fuzzy β-covering $\mathcal{C}$ is more significant for the decision D.

Based on the above definitions, a forward greedy algorithm for fuzzy β-covering reduction in an MFβCDT is proposed as shown in Algorithm 1, and its time complexity is also discussed.

As described above, Algorithm 1 uses the monotone conditional entropy as an index of significance of a fuzzy β-covering and takes $\mathrm{SIG}^{\beta,\lambda}(\mathcal{C},\mathcal{G},D)$ as the termination condition. When adding any remaining fuzzy β-covering does not decrease the monotonic conditional entropy of D, the algorithm stops. For an MFβCDT with m samples and n fuzzy β-coverings, the time complexity of $O(n*m^2)$ is needed to compute the fuzzy β-covering relation, and the worst search time of $O(n(n+1)/2*m^2)$ is needed to find a reduct. In summary, the overall time complexity of Algorithm 1 is $O(n(n+1)/2*m^2)$.

In order to verify the effectiveness and rationality of our proposed algorithm, an example is given as follows.

Example 5: (Following Example 2) Given an MFβCDT $<U,\mathcal{M},D>$, where $U=\{x_1,x_2,x_3,x_4\}$, $\mathcal{M}=\{\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_3,\mathcal{C}_4\}$ is a family of fuzzy β-coverings of U, $\mathcal{C}_j=\{C_1^j,C_2^j,C_3^j,C_4^j\}$, and $\mathcal{C}_j\in\mathcal{M}$. $U/D=\{\{x_1,x_2\},\{x_3\},\{x_4\}\}$ is a partition of U on D.

Suppose that $R_{\mathcal{C}}^{\beta}={}^1\mathrm{N}_{\mathcal{C}}^{\beta}$, where $\mathcal{T}=\mathcal{T}_M$ and $\mathcal{I}=\mathcal{I}_L$. Let $\beta=0.5$ and $\lambda=0.1$, then we can find a reduct red of $\mathcal{M}$ as follows.

According to Definition 17, we can obtain that $H^{\beta,\lambda}(D\mid\mathcal{C}_1)=0.26$, $H^{\beta,\lambda}(D\mid\mathcal{C}_2)=0.30$, $H^{\beta,\lambda}(D\mid\mathcal{C}_3)=0.37$, and $H^{\beta,\lambda}(D\mid\mathcal{C}_4)=0.23$. Consequently, we have $\mathrm{red}=\{\mathcal{C}_4\}$. Since $\mathcal{G}=\{\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_3\}\neq\varnothing$ and $\mathrm{SIG}(\mathcal{C}_4,\varnothing,D)=2.17>0$, we need to continue the next step. We calculate that $H^{\beta,\lambda}(D\mid\{\mathcal{C}_1,\mathcal{C}_4\})=0$, $H^{\beta,\lambda}(D\mid\{\mathcal{C}_2,\mathcal{C}_4\})=0$, and $H^{\beta,\lambda}(D\mid\{\mathcal{C}_3,\mathcal{C}_4\})=0$. Subsequently, we obtain a new reduct $\mathrm{red}=\{\mathcal{C}_1,\mathcal{C}_4\}$. Since $G=\{\mathcal{C}_2,\mathcal{C}_3\}\neq\varnothing$ and $\mathrm{SIG}(\mathcal{C}_1,\{\mathcal{C}_4\},D)=0.23>0$, we need further calculations. Similarly, we have $H^{\beta,\lambda}(D\mid\{\mathcal{C}_1,\mathcal{C}_2,\mathcal{C}_4\})=0$ and $H^{\beta,\lambda}(D\mid\{\mathcal{C}_1,\mathcal{C}_3,\mathcal{C}_4\})=0$. Thus, we can obtain the final reduct $\mathrm{red}=\{\mathcal{C}_1,\mathcal{C}_4\}$ with $\mathrm{SIG}(\mathcal{C}_2,\{\mathcal{C}_1,\mathcal{C}_4\},D)=0$ and $\mathrm{SIG}(\mathcal{C}_3,\{\mathcal{C}_1,\mathcal{C}_4\},D)=0$.

Furthermore, the connections and differences between the proposed method and the existing methods of attribute reduction based on fuzzy β-covering theory are also given as follows.

According to [26], we can determine that both the fuzzy β-covering reduction method proposed in this article and the fuzzy β-covering reduction method proposed in [26] adopt information theory to measure the significance of fuzzy β-coverings. Moreover, the attribute reduction framework proposed in this article and the attribute reduction method proposed in [26] are constructed based on the fuzzy β-covering reduction method. Further, there are some differences between them. First, the attribute reduction method in [26] simply uses a given formula to construct fuzzy β-coverings manually, whereas the method proposed in this article uses fuzzy information granulation, which is more general. Second, the fuzzy β-covering reduction method in [26] may have a stability problem due to the nonmonotonicity of the conditional discrimination measure, while the fuzzy β-covering reduction method proposed in this article does not have this problem.

Remark 3: Compared with the method in [26], the method proposed in this article has some advantages. In this article, we propose a general framework for attribute reduction based on fuzzy β-covering reduction, which has stronger generalization performance. Moreover, our proposed fuzzy β-covering reduction method adopts the monotonic conditional information measure, so it has better stability.

VI. Experimental Analysis

In this section, we compare four state-of-the-art attribute reduction algorithms with our proposed algorithm (MFBC) experimentally. These four algorithms are the algorithm based on the conditional discernibility measure of fuzzy βcovering (FBC) [26], the algorithm based on neighborhood discrimination index (HANDI) [32], the algorithm based on the conditional entropy of fuzzy similarity relation (FSRBCE) [31], and the algorithm-based fuzzy rough sets (FRS) [6], [38] respectively. One algorithm (FRS) is based on the dependency degree, and the other three are based on the theory of information entropy. The proposed algorithm is evaluated from the following four aspects:

1) size of selected attribute subset;

2) classification performance of selected attribute subset;

3) correlation between five algorithms, that is, statistical analysis;

4) robustness of the proposed algorithm, that is, parameter analysis.

All experiments are implemented by MATLAB R2019a and run on a personal computer with an Intel Core i7-6700HQ CPU at 2.60 GHz, 32.0 GB RAM, and 64-bit Windows 10.

TABLE II
DESCRIPTION OF DATASETS

No.	Dataset	Samples	Attributes	Classes
1	Glass	214	9	6
2	Wine	178	13	3
3	Ionosphere	351	34	2
4	Soybean	307	35	19
5	Promoters	106	57	2
6	MuskV1	476	166	2
7	German	1000	24	2
8	ColonTumor	62	2000	2
9	DLBCLOutcome	58	7129	2
10	DLBCLTumor	77	7129	2
11	Ovarian61902	253	15154	2
12	Breast	97	24481	2

A. Experimental Design

In our experiments, 12 open datasets are collected as benchmarks for evaluating algorithm performance. Among them, the first seven low-dimensional datasets are obtained from UCI Machine Learning Repository [39], and the other high-dimensional datasets are downloaded from KRBD Repository [40]. The details of these datasets are given in Table II. Moreover, before attribute reduction, all data are normalized to the interval [0, 1].

Two classical classifiers, the decision tree (CART), and the K-nearest neighbor (KNN, $K = 3$), are used to obtain the classification accuracies of all datasets. Classification and regression trees, shortly CART, is a typical binary decision tree. CART algorithm uses Gini coefficient to replace the information gain ratio in C4.5 algorithm to select attributes. Gini coefficient represents the impure of the model. The smaller the Gini coefficient is, the lower the impure is, and the better the attribute subset is. K-Nearest Neighbor, shortly KNN, is one of the simplest methods in data mining classification technology. The core idea of KNN algorithm is that each sample can be represented by its closest K adjacent values. In these experiments, we use two functions $fitctree$ and $fitcknn$ provided by the statistics and machine learning toolbox of matlab to complete the classification task. Besides, the tenfold cross validation technology is also employed to obtain the classification accuracies of all datasets. Furthermore, the mean and the standard deviation format of classification accuracies are shown as the final performance.

According to Algorithm 1, we can see that two parameters $\beta$and λ may have an impact on the results of attribute reduction of our proposed algorithm. Thus, we select an optimal attribute subset for each dataset by adjusting the value of the covering threshold $\beta$from 0.5 to 0.9 with a step of 0.1 and the value of the neighborhood radius λ from 0.1 to 0.5 with a step of 0.1. Further, we can see how these two parameters affect our proposed algorithm. In order to fairly compare our proposed algorithm with the other four ones, we set $\beta$to a value between 0.5 and 0.9 in steps of 0.1 and δ to a value between 0.1 and 0.5 in steps of 0.1 for FBC, and set ε to a value between 0.1 and 0.5 in steps of 0.1 for HANDI. All experimental results are displayed with the highest classification accuracy as follows. All attribute reduction algorithms adopt the sequential forward search strategy in our experiments.

B. Analysis of the Experimental Results

In this section, the experimental results based on the four evaluation indexes mentioned above are respectively shown in the following. The experimental results verify the superiority and stability of our proposed algorithm.

Table III lists the sizes of attribute subsets selected by five attribute reduction algorithms and original data on 12 datasets. In particular, three algorithms (HANDI, FBC, and MFBC) show the sizes of attribute subsets with the highest classification accuracies under different classifiers for each dataset. For HANDI, the value in parentheses represents the value of the parameter ε with the highest classification accuracy of a dataset under a classifier. Similarly, the values in parentheses represent the optimal parameter values for FBC (β/δ) and MFBC (β/λ). According to the results given in Table III, we can determine that these five algorithms can effectively select attributes. The smallest average size of attribute subsets reaches only 5.08 obtained by FSRBCE, and the largest attains only 17 obtained by FRS. For MFBC, the average dimensional reduction rate under two classifiers is as high as 99.82% for 12 datasets.

Tables IV and V display, respectively, the classification performance of five attribute reduction algorithms under two classifiers, where the underlined data represent the highest classification accuracy obtained by these five algorithms under the same dataset. The parameter settings of HANDI, FBC, and MFBC for obtaining the classification accuracies in Tables IV and V can be found in Table III. From Tables IV and V, we can clearly see that the classification accuracy of MFBC outperforms the other four algorithms on most of datasets. Out of the total 24 cases, MFBC algorithm obtains the highest classification accuracy in 14 cases, and the FBC and HANDI algorithms acquire the highest classification accuracies in 6 and 4 cases, respectively. FSRBCE algorithm achieves it only one case, and FRS algorithm attains it in two cases. Further, MFBC algorithm improves the average classification accuracy 7.2% for CART and 9.14% for KNN on the basis of the raw data. Thus, the average classification accuracy of MFBC algorithm is higher than other four ones under these two classifiers. In general, our proposed algorithm is superior to other four ones in classification performance.

Besides, Friedman [41] and Bonferroni–Dunn tests [42] are employed to analyze the significance of our proposed algorithm from the viewpoint of statistics. Suppose that m and k are the numbers of datasets and algorithms, respectively. r_j is the average rank of the jth algorithm for m datasets. Then, the formula of Friedman statistic under null-hypothesis is given as

$$F_F = \frac{(m-1)\chi_F^2}{m(k-1) - \chi_F^2} \tag{26}$$

where F_F follows a Fisher distribution with $k-1$ and $(k-1)(m-1)$ degrees of freedom, and

$$\chi_F^2 = \frac{12m}{k(k+1)} \left(\sum_{j=1}^{k} r_j^2 - \frac{k(k+1)^2}{4} \right).$$

Especially, the critical value is $F_F(4, 44) = 2.08$ at the significance level $\alpha = 0.1$.

Based on the above mentioned, we can easily obtain the ranks of classification accuracies of five algorithms under two classifiers, given in Tables VI and VII, respectively. According to (26), we can obtain that the value of F_F under CART is 5.49 and the value of F_F under KNN is 2.21. Further, we can see that the values of test statistic for CART and KNN classifiers are larger than the critical value, which implies that the null-hypothesis at $\alpha = 0.1$ is rejected. Thus, we can determine that these five attribute reduction algorithms are significantly different.

TABLE III
AVERAGE SIZES OF ATTRIBUTE SUBSETS

<table>
<tr><th rowspan="2">Dataset</th><th rowspan="2">ALL</th><th rowspan="2">FRS</th><th rowspan="2">FSRBCE</th><th colspan="2">HANDI</th><th colspan="2">FBC</th><th colspan="2">MFBC</th></tr>
<tr><th>CART</th><th>KNN</th><th>CART</th><th>KNN</th><th>CART</th><th>KNN</th></tr>
<tr><td>Glass</td><td>9</td><td>5</td><td>3</td><td>7(0.1)</td><td>5(0.2)</td><td>6(0.6/0.5)</td><td>8(0.7/0.1)</td><td colspan="2">9(0.5/0.1)</td></tr>
<tr><td>Wine</td><td>13</td><td>7</td><td>3</td><td colspan="2">12(0.1)</td><td>6(0.7/0.1)</td><td>7(0.9/0.1)</td><td>6 (0.5/0.4)</td><td>7(0.5/0.3)</td></tr>
<tr><td>Ionosphere</td><td>34</td><td>9</td><td>4</td><td>13(0.3)</td><td>6(0.1)</td><td>4(0.7/0.3)</td><td>4(0.6/0.1)</td><td>15(0.5/0.2)</td><td>11(0.5/0.4)</td></tr>
<tr><td>Soybean</td><td>35</td><td>24</td><td>9</td><td colspan="2">30(0.1)</td><td colspan="2">12(0.5/0.1)</td><td colspan="2">16(0.5/0.1)</td></tr>
<tr><td>Promoters</td><td>57</td><td>30</td><td>2</td><td colspan="2">12(0.1)</td><td colspan="2">4(0.5/0.1)</td><td>6(0.5/0.1)</td><td>4(0.5/0.3)</td></tr>
<tr><td>MuskV1</td><td>166</td><td>19</td><td>6</td><td>8(0.1)</td><td>22(0.4)</td><td>7(0.5/0.2)</td><td>8(0.5/0.1)</td><td>12(0.5/0.4)</td><td>14 (0.5/0.1)</td></tr>
<tr><td>German</td><td>24</td><td>24</td><td>1</td><td colspan="2">21(0.1)</td><td>14(0.5/0.1)</td><td>7(0.5/0.5)</td><td>16(0.5/0.2)</td><td>19(0.5/0.1)</td></tr>
<tr><td>ColonTumor</td><td>2000</td><td>17</td><td>7</td><td>4(0.1)</td><td>9(0.5)</td><td>3(0.8/0.5)</td><td>4(0.6/0.1)</td><td>5(0.5/0.4)</td><td>5(0.5/0.2)</td></tr>
<tr><td>DLBCLOutcome</td><td>7129</td><td>23</td><td>7</td><td colspan="2">7(0.4)</td><td>4(0.8/0.4)</td><td>4(0.7/0.1)</td><td colspan="2">5(0.5/0.2)</td></tr>
<tr><td>DLBCLTumor</td><td>7129</td><td>15</td><td>6</td><td>3(0.1)</td><td>7(0.5)</td><td colspan="2">4(0.9/0.2)</td><td colspan="2">4(0.5/0.4)</td></tr>
<tr><td>Ovarian61902</td><td>15154</td><td>8</td><td>2</td><td colspan="2">3(0.3)</td><td>2(0.5/0.3)</td><td>2(0.5/0.1)</td><td colspan="2">4(0.5/0.1)</td></tr>
<tr><td>Breast</td><td>24481</td><td>23</td><td>11</td><td>6(0.3)</td><td>10(0.5)</td><td>3(0.5/0.3)</td><td>5(0.8/0.1)</td><td colspan="2">5(0.5/0.3)</td></tr>
<tr><td>Average</td><td>4685.92</td><td>17</td><td>5.08</td><td>10.50</td><td>12</td><td>5.75</td><td>5.75</td><td>8.58</td><td>8.58</td></tr>
</table>

TABLE IV
CLASSIFICATION ACCURACIES OF ATTRIBUTE SUBSETS WITH CART (%)

Dataset	ALL	FRS	FSRBCE	HANDI	FBC	MFBC
Glass	69.26±10.04	60.04±9.24	59.38±9.14	62.13±6.35	67.46±11.04	69.24±8.83
Wine	91.56±5.51	93.88±6.56	94.43±5.75	92.15±6.53	94.96±4.89	94.99±4.02
Ionosphere	87.70±5.97	88.03±3.77	90.56±5.23	91.13±5.66	90.59±4.21	91.72±4.59
Soybean	90.47±3.27	89.61±3.89	74.36±4.57	89.45±4.28	83.89±5.20	85.09±3.85
Promoters	76.12±11.28	79.77±11.40	83.77±10.03	78.61±13.92	84.94±6.14	85.27±11.30
MuskV1	79.45±4.25	78.80±7.28	73.51±5.93	80.02±6.51	81.50±6.40	79.39±7.63
German	71.70±4.92	72.40±5.21	70.00±0.00	72.90±4.65	71.80±4.64	72.10±4.20
ColonTumor	72.38±15.92	77.62±10.89	82.38±11.78	82.62±13.00	84.05±16.15	88.57±7.92
DLBCLOutcome	53.81±21.58	71.29±28.04	63.71±17.01	70.14±12.58	77.14±19.22	80.24±14.31
DLBCLTumor	84.46±11.85	91.96±9.68	95.00±6.46	93.57±6.80	88.75±16.08	92.50±8.74
Ovarian61902	98.83±1.89	98.40±2.07	98.03±2.81	98.40±2.80	98.80±1.93	98.40±2.80
Breast	53.70±15.85	68.48±20.51	64.34±13.17	72.87±12.87	78.33±19.08	79.50±9.92
Average	77.45±9.36	80.86±9.88	79.12±7.66	82.00±8.00	83.52±9.58	84.75±7.34

TABLE V
CLASSIFICATION ACCURACIES OF ATTRIBUTE SUBSETS WITH *K*NN (%)

Dataset	ALL	FRS	FSRBCE	HANDI	FBC	MFBC
Glass	73.42±8.73	67.44±8.10	63.51±8.48	67.44±8.10	72.10±7.95	73.42±8.73
Wine	94.34±3.82	94.99±4.88	94.43±5.04	93.26±5.81	98.86±2.41	98.86±2.41
Ionosphere	84.90±4.67	90.60±2.33	91.15±3.47	90.31±3.87	92.56±5.91	88.89±3.92
Soybean	90.76±4.33	87.09±6.14	71.87±5.54	89.89±4.34	76.71±4.07	84.31±4.80
Promoters	68.80±15.80	77.55±14.97	78.38±11.30	78.03±15.83	85.76±13.94	85.76±13.94
MuskV1	88.88±4.59	82.61±6.00	78.38±5.04	85.93±4.03	83.83±4.69	86.13±3.73
German	71.50±4.55	71.50±4.55	64.20±3.85	71.30±4.19	70.50±2.95	72.40±5.27
ColonTumor	72.62±15.07	89.05±10.18	87.14±12.41	90.48±8.25	87.38±14.64	81.19±19.10
DLBCLOutcome	45.81±25.30	66.14±16.06	79.48±15.69	82.24±13.49	85.67±14.49	82.48±18.02
DLBCLTumor	86.79±14.37	98.57±4.52	97.50±5.27	97.32±5.66	93.57±9.00	93.57±9.00
Ovarian61902	93.29±2.66	99.63±1.17	99.60±1.26	100.00±0.00	99.60±1.26	100.00±0.00
Breast	53.50±19.45	77.63±13.06	76.18±10.24	81.42±9.51	79.11±10.85	87.33±10.94
Average	77.05±10.28	83.57±7.66	81.82±7.30	85.64±6.92	85.47±7.68	86.19±8.32

TABLE VI
RANK OF FIVE ATTRIBUTE REDUCTION ALGORITHMS WITH CART

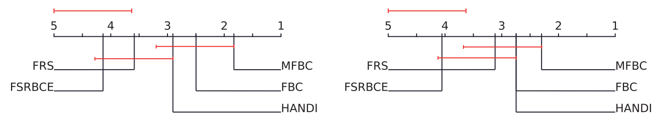
Dataset	FRS	FSRBCE	HANDI	FBC	MFBC
Glass	4	5	3	2	1
Wine	4	3	5	2	1
Ionosphere	5	4	2	3	1
Soybean	1	5	2	4	3
Promoters	4	3	5	2	1
MuskV1	4	5	2	1	3
German	2	5	1	4	3
ColonTumor	5	4	3	2	1
DLBCLOutcome	3	5	4	2	1
DLBCLTumor	4	1	2	5	3
Ovarian61902	3	5	3	1	3
Breast	4	5	3	2	1
Average	3.58	4.17	2.92	2.50	1.83

TABLE VII
RANK OF FIVE ATTRIBUTE REDUCTION ALGORITHMS WITH *K*NN

Dataset	FRS	FSRBCE	HANDI	FBC	MFBC
Glass	3.5	5	3.5	2	1
Wine	3	4	5	1.5	1.5
Ionosphere	3	2	4	1	5
Soybean	2	5	1	4	3
Promoters	5	3	4	1.5	1.5
MuskV1	4	5	2	3	1
German	2	5	3	4	1
ColonTumor	2	4	1	3	5
DLBCLOutcome	5	4	3	1	2
DLBCLTumor	1	2	3	4.5	4.5
Ovarian61902	3	4.5	1.5	4.5	1.5
Breast	4	5	2	3	1
Average	3.13	4.04	2.75	2.75	2.33

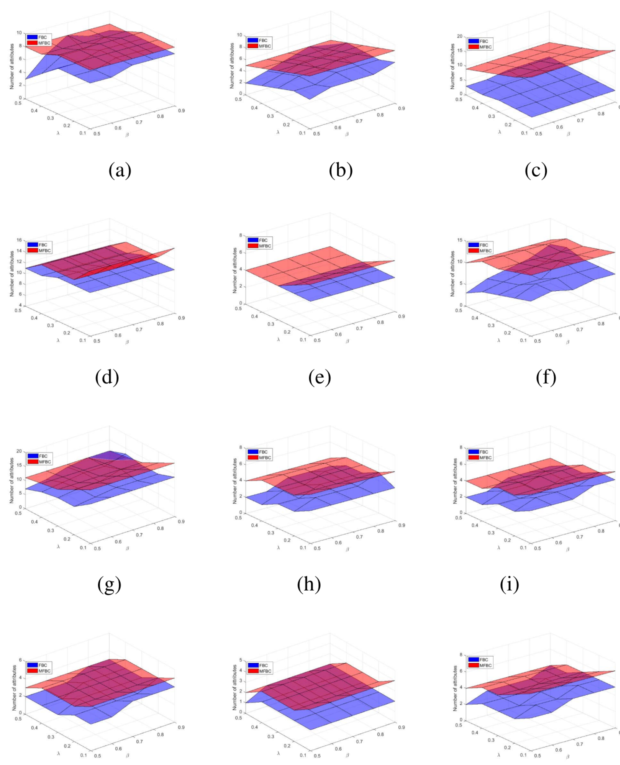
Therefore, Bonferroni–Dunn test is needed to further explore the difference between these five algorithms in classification performance. The performance of two algorithms is regarded as different if their average rank distance exceeds the critical distance

$$\mathrm{CD}_{\alpha} = q_{\alpha}\sqrt{\frac{k(k+1)}{6m}} \tag{27}$$



(a) (b)

Fig. 3. Accuracy comparison with five algorithms. (a) CART (b) *K*NN.



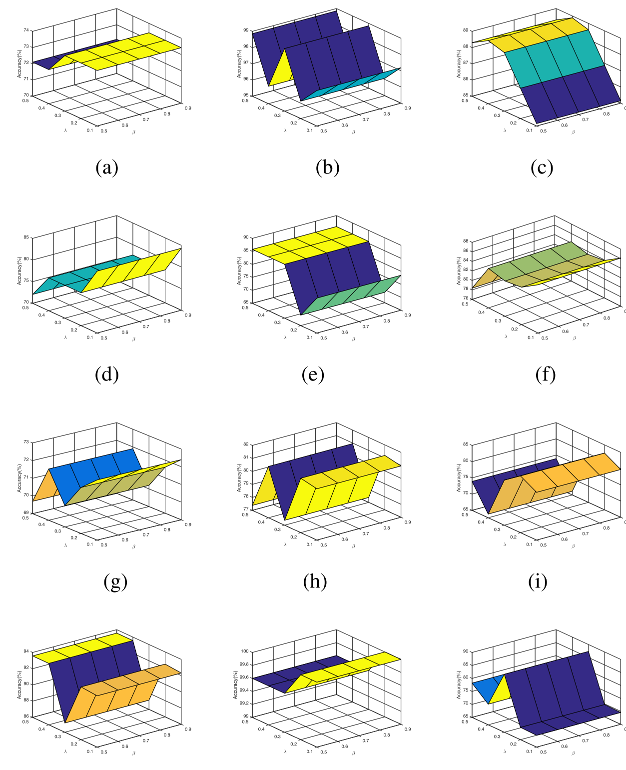
(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

Fig. 4. Number varying of selected attributes with $\beta$and λ. (a) Glass. (b) Wine. (c) Ionosphere. (d) Soybean. (e) Promoters. (f) MuskV1. (g) German. (h) ColonTumor. (i) DLBCLOutcome. (j) DLBCLTumor. (k) Ovarian61902. (l) Breast.

where the critical value $q_{\alpha} = 2.128$ at the significance level $\alpha = 0.1$ can be obtained in [43].

By (27), we can attain $\mathrm{CD}_{0.1} = 1.37$ when $m = 12$ and $k = 5$. In order to more intuitively judge the performance difference between these five algorithms, the critical distance diagrams for two classifiers are shown as Fig. 3, where the red line denotes the critical distance. Especially, if MFBC algorithm and another are both within the red line, then they are not significantly different; otherwise, they are significantly different. These two tests demonstrate that MFBC algorithm is statistically better than other four ones for CART and *K*NN classifiers.

In order to further compare the difference between the MFBC and FBC algorithms, the diagrams of sizes of attribute subsets selected by these two algorithms with the same parameter changes on 12 datasets are shown in Fig. 4. From Fig. 4, we can observe that attribute subsets selected by these two algorithms are different for 12 datasets, and the number of attributes selected by FBC algorithm is less than that of MFBC algorithm on all datasets. Fortunately, the difference of sizes of attribute subsets selected by these two algorithms is very small. The reason is that



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k) (l)

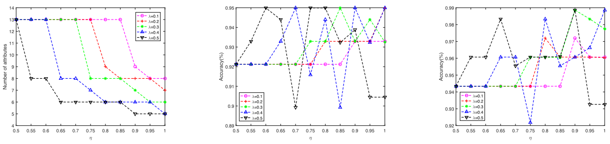
Fig. 5. Accuracy varying of datasets with $\beta$and λ. (a) Glass. (b) Wine. (c) Ionosphere. (d) Soybean. (e) Promoters. (f) MuskV1. (g) German. (h) ColonTumor. (i) DLBCLOutcome. (j) DLBCLTumor. (k) Ovarian61902. (l) Breast.

these two algorithms use different but similar measure methods to select attributes.

Finally, we conduct a series of experiments to verify the robustness of our proposed algorithm. From Algorithm 1, we can determine that the covering threshold $\beta$and the neighborhood radius λ may affect the performance of our proposed algorithm. To this end, we implemented a variation on the classification accuracy by adjusting the value of $\beta$to vary from 0.5 to 0.9 with a step of 0.1 and λ to vary from 0.1 to 0.5 with a step of 0.1. The classification accuracies of datasets varying with $\beta$and λ on *K*NN are shown in Fig. 5. The experimental results obtained by CART are roughly consistent with *K*NN.

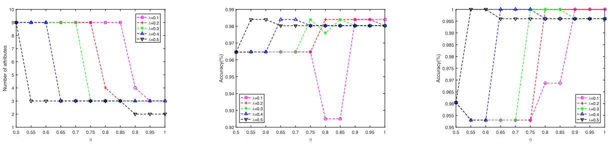
From Fig. 5, we can see that the classification accuracy of dataset remains constant as the value of $\beta$is changed when λ is a constant value for 12 datasets. It means that the covering threshold βcan be ignored when analyzing the neighborhood radius λ. The classification accuracy of dataset decreases with the increasing value of λ on most of datasets from the overall trend of change. Furthermore, we observe that most of datasets have achieved high classification accuracy. According to Figs. 4 and 5, we can ensure that our proposed algorithm is robust and stable.

In addition, we can see that the proposed framework for attribute reduction consists of two parts from Section V: the acquisition of MFβCDTs and the fuzzy β-covering reduction. In the above, we have analyzed the effects of the parameters λ and βin the part of fuzzy β-covering reduction on the experimental results (the number of attributes selected and the classification accuracy). Next, we take a low-dimensional dataset Wine and a



(a) (b) (c)

Fig. 6. Result varying of the dataset Wine with η. (a) MFBC. (b) CART. (c) *K*NN.



(a) (b) (c)

Fig. 7. Result varying of the dataset Ovarian61902 with η. (a) MFBC. (b) CART. (c) *K*NN.

high-dimensional dataset Ovarian61902 as examples to analyze the parameter βin the part of acquisition of MFβCDTs. Through the previous experimental analysis, we can determine that the proposed method is not sensitive to the parameter β, i.e., the parameter βhas little effect on the experimental results. Therefore, we only analyze the effects of the parameter η change on the experimental results when the parameter λ is varied and the parameter $\beta$is fixed. The parameter λ is set to a value between 0.1 and 0.5 in steps of 0.1, the parameter $\beta$is set to 0.5, and the parameter η is set to a value between 0.5 and 1 in steps of 0.05.

Figs. 6 and 7 show the results obtained from the analysis of the parameter η on the datasets Wine and Ovarian61902, respectively. The first plot shows the results for the number of selected attributes as the parameter η varies, the second plot shows the results for the classification accuracies of selected attribute subsets as the parameter η varies under the CART classifier, and the third plot shows the results for the classification accuracies of selected attribute subsets as the parameter η varies under the *K*NN classifier. From Figs. 6 and 7, we can see that the number of selected attributes is almost always the lowest and the classification accuracy is almost always the highest when $\eta = 0.9$. This conclusion verifies that the value of η recommended in the literature [26], [27] is set at 0.9 is reasonable.

In summary, the experimental results show that MFBC algorithm is indeed better than other four ones for CART and *K*NN classifiers.

VII. Conclusion

In order to solve the problem that fuzzy β-covering is difficult to find directly from the real data, a general approach of obtaining fuzzy β-covering is given, i.e., fuzzy information granulation. Since fuzzy information granulation is capable of generating multiple fuzzy β-coverings, the concept of MFβCASs is proposed. Then, several information measures are investigated in the context of MFβCASs. Moreover, a novel heuristic fuzzy β-covering reduction algorithm and a general framework of attribute reduction are also designed. Finally, a series of experiments verify the effectiveness and superiority of our proposed method. The main contributions of this article are listed as follows.

1) To apply the relevant theory of fuzzy β-covering to solve some practical problems, we introduce the concept of MFβCASs and present its related theories.
2) In the context of MFβCASs, we investigate several new information measures for characterizing the distinguishing ability of fuzzy β-covering family.
3) Based on the measure of monotone conditional entropy, we propose a novel heuristic fuzzy β-covering reduction method in an MFβCDT.
4) By virtue of fuzzy information granulation, we design a general framework of attribute reduction base on fuzzy β-covering reduction.

The concept of MFβCASs not only enriches the theory of fuzzy β-covering but also expands its application. Theoretically, FβCASs are a special case of MFβCASs, so we can continue to study the related theory of fuzzy β-covering on the basis of FβCASs, that is, MFβCASs are compatible with the existing fuzzy β-covering theory in the context of FβCASs. Moreover, the new fuzzy β-covering theory can further be developed in the context of MFβCASs. In terms of applications, MFβCASs are a concept derived to solve the problem that fuzzy β-covering is difficult to obtain. It allows the fuzzy β-covering theory to be applied to solve some more complex practical problems. Since the data stored in MFβCASs have been processed into the effective classification information, the running time of attribute reduction method in the context of MFβCASs is reduced. However, the concept of MFβCASs has some limitations. First, MFβCASs require the valid analysis and processing of the real data before it can be obtained. Second, the methods in the context of MFβCASs are not suitable for large-scale high-dimensional datasets due to the large amount of storage space required by MFβCASs.

References

[1] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, 1982.

[2] W. Zakowski, "Approximations in the space (u, π)," *Demonstratio Math.*, vol. 16, no. 3, pp. 761–770, 1983.

[3] Y. Yao and B. Yao, "Covering-based rough set approximations," *Inf. Sci.*, vol. 200, pp. 91–107, 2012.

[4] J. Chen, Y. Lin, G. Lin, J. Li, and Y. Zhang, "Attribute reduction of covering decision systems by hypergraph model," *Knowl.-Based Syst.*, vol. 118, pp. 93–104, 2017.

[5] Z. Li, D. Huang, X. Liu, N. Xie, and G. Zhang, "Information structures in a covering information system," *Inf. Sci.*, vol. 507, pp. 449–471, 2020.

[6] D. Dubois and H. Prade, "Rough fuzzy sets and fuzzy rough sets," *Int. J. Gen. Syst.*, vol. 17, no. 2/3, pp. 191–209, 1990.

[7] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 1, pp. 338–353, 1965.

[8] D. Chen, L. Zhang, S. Zhao, Q. Hu, and P. Zhu, "A novel algorithm for finding reducts with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 2, pp. 385–389, Apr. 2012.

[9] C. Wang et al., "A fitting model for feature selection with fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 4, pp. 741–753, Aug. 2017.

[10] S. Zhao et al., "An accelerator for rule induction in fuzzy rough theory," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 12, pp. 3635–3649, Dec. 2021.

[11] T. Deng, Y. Chen, W. Xu, and Q. Dai, "A novel approach to fuzzy rough sets based on a fuzzy covering," *Inf. Sci.*, vol. 177, no. 11, pp. 2308–2326, 2007.

[12] T. Li, Y. Leung, and W. Zhang, "Generalized fuzzy rough approximation operators based on fuzzy coverings," *Int. J. Approx. Reason.*, vol. 48, no. 3, pp. 836–856, 2008.

[13] T. Feng, S. Zhang, and J. Mi, "The reduction and fusion of fuzzy covering systems based on the evidence theory," *Int. J. Approx. Reason.*, vol. 53, no. 1, pp. 87–103, 2012.

[14] L. D'eer, C. Cornelis, and L. Godo, "Fuzzy neighborhood operators based on fuzzy coverings," *Fuzzy Sets Syst.*, vol. 312, pp. 17–35, 2017.

[15] L. Ma, "Two fuzzy covering rough set models and their generalizations over fuzzy lattices," *Fuzzy Sets Syst.*, vol. 294, pp. 1–17, 2016.
[16] B. Huang, C. Guo, H. Li, G. Feng, and X. Zhou, "An intuitionistic fuzzy graded covering rough set," *Knowl.-Based Syst.*, vol. 107, pp. 155–178, 2016.
[17] B. Yang and B. Hu, "Fuzzy neighborhood operators and derived fuzzy coverings," *Fuzzy Sets Syst.*, vol. 370, pp. 1–33, 2019.
[18] X. Zhang and J. Wang, "Fuzzy β-covering approximation spaces," *Int. J. Approx. Reason.*, vol. 126, pp. 27–47, 2020.
[19] H. Jiang, J. Zhan, and D. Chen, "Covering-based variable precision (I, T)-fuzzy rough sets with applications to multiattribute decision-making," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 8, pp. 1558–1572, Aug. 2019.
[20] X. Zhang, J. Wang, J. Zhan, and J. Dai, "Fuzzy measures and Choquet integrals based on fuzzy covering rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 7, pp. 2360–2374, Jul. 2022.
[21] C. Cornelis, M. De Cock, and E. E. Kerre, "Intuitionistic fuzzy rough sets: At the crossroads of imperfect knowledge," *Expert Syst.*, vol. 20, no. 5, pp. 260–270, 2003.
[22] A. Mariello and R. Battiti, "Feature selection based on the neighborhood entropy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 12, pp. 6313–6322, Dec. 2018.
[23] W. Ding, C.-T. Lin, and W. Pedrycz, "Multiple relevant feature ensemble selection based on multilayer co-evolutionary consensus mapreduce," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 425–439, Feb. 2020.
[24] T. Yang, X. Zhong, G. Lang, Y. Qian, and J. Dai, "Granular matrix: A new approach for granular structure reduction and redundancy evaluation," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 12, pp. 3133–3144, Dec. 2020.
[25] Z. Huang and J. Li, "A fitting model for attribute reduction with fuzzy β-covering," *Fuzzy Sets Syst.*, vol. 413, pp. 114–137, 2021.
[26] Z. Huang and J. Li, "Discernibility measures for fuzzy β covering and their application," *IEEE Trans. Cybern.*, early access, Mar. 9, 2021, doi: 10.1109/TCYB.2021.3054742.
[27] Z. Huang, J. Li, and Y. Qian, "Noise-tolerant fuzzy covering-based multigranulation rough sets and feature subset selection," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 7, pp. 2721–2735, Jul. 2022.
[28] K. Zhang, J. Zhan, and W. Wu, "On multi-criteria decision-making method based on a fuzzy rough set model with fuzzy α-neighborhoods," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 9, pp. 2491–2505, Sep. 2021.
[29] B. Yang and B. Hu, "A fuzzy covering-based rough set model and its generalization over fuzzy lattice," *Inf. Sci.*, vol. 367, pp. 463–486, 2016.
[30] K. Zhang and J. Dai, "Three-way multi-criteria group decision-making method in a fuzzy β-covering group approximation space," *Inf. Sci.*, vol. 599, pp. 1–24, 2022.
[31] Q. Hu, D. Yu, Z. Xie, and J. Liu, "Fuzzy probabilistic approximation spaces and their information measures," *IEEE Trans. Fuzzy Syst.*, vol. 14, no. 2, pp. 191–201, Apr. 2006.
[32] C. Wang, Q. Hu, X. Wang, D. Chen, Y. Qian, and Z. Dong, "Feature selection based on neighborhood discrimination index," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2986–2999, Jul. 2018.
[33] Q. Hu, Z. Xie, and D. Yu, "Comments on 'fuzzy probabilistic approximation spaces and their information measures'," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 2, pp. 549–551, Apr. 2008.
[34] J. Dai, W. Wang, and Q. Xu, "An uncertainty measure for incomplete decision tables and its applications," *IEEE Trans. Cybern.*, vol. 43, no. 4, pp. 1277–1289, Aug. 2013.
[35] L. A. Zadeh, "Fuzzy logic= computing with words," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 103–111, May 1996.
[36] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, 1997.
[37] Z. Yuan, H. Chen, P. Xie, P. Zhang, J. Liu, and T. Li, "Attribute reduction methods in fuzzy rough set theory: An overview, comparative experiments, and new directions," *Appl. Soft Comput.*, vol. 107, 2021, Art. no. 107353.
[38] R. Jensen and Q. Shen, "Fuzzy-rough sets assisted attribute selection," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 1, pp. 73–89, Feb. 2007.
[39] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml
[40] J. Li and H. Liu, "Kent Ridge bio-medical data set repository," 2002. [Online]. Available: http://datam.i2r.a-star.edu.sg/datasets/krbd/index.html
[41] M. Friedman, "A comparison of alternative tests of significance for the problem of M rankings," *Ann. Math. Statist.*, vol. 11, no. 1, pp. 86–92, 1940.
[42] O. J. Dunn, "Multiple comparisons among means," *J. Amer. Statist. Assoc.*, vol. 56, no. 293, pp. 52–64, 1961.
[43] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, 2006.