This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems

Noise-tolerant fuzzy β covering based multigranulation rough sets and feature subset selection

Zhehuang Huang, Jinjin Li, and Yuhua Qian*, *Member, IEEE*

***Abstract*—As a novel fuzzy covering, fuzzy β covering has attracted considerable attention. However, traditional fuzzy β covering based rough sets and most of its extended models can not well fit the distribution of samples in real data, which limits their application in classification learning and decision making. First, the upper and lower approximations of these models have no inclusion relation, so they can not characterize a given objective concept accurately. Moreover, most of these models are hard to resist the influence of noise data, resulting in poor robustness in feature learning. For these reasons, a robust rough set model is set forth by combining fuzzy rough sets, covering based rough sets, and multigranulation rough sets. To this end, the optimistic and pessimistic lower and upper approximations of a target concept is reconstructed by means of the fuzzy β neighborhood related to a family of fuzzy coverings, and a new multigranulation fuzzy rough set model is presented. Furthermore, fuzzy dependency function is introduced to evaluate the classification ability of a family of fuzzy β coverings at different granularity level. The dimensionality reduction of a given fuzzy covering decision table is carried out from the perspective of maintaining the discrimination power, and a forward algorithm for feature selection is developed by using the optimistic significance of candidate features as heuristic information. Three groups of numerical experiments on 16 different types of data sets demonstrate that the proposed model exhibits good robustness on data sets contaminated with noise, and outperforms some state-of-the-art feature learning algorithms in terms of classification accuracy and the size of selected feature subset.**

***Index Terms*—Fuzzy β covering, multigranulation rough sets, covering rough sets, feature selection.**

I. Introduction

FEATURE selection is an effective technique for knowledge reduction in the fields of machine learning and granular computing [1]-[4]. By reducing redundant features, it can improve the generalization ability of learning model, and simplify the complexity of computation [5]-[7].

This work was supported by the National Natural Science Foundation of China (No:11871259, 11701258), the Natural Science Foundation of Fujian Province (2017J01114) and the High-level Talents Start-up Project of Huaqiao University (16BS814). (Corresponding author: Yuhua Qian.)

Z. H. Huang is with School of Mathematics Sciences, Huaqiao University, Quanzhou, 362021, Fujian, China (e-mail: startstart1@163.com).

J.J. Li is with School of Mathematics Sciences and Statistics, Minnan Normal University, ZhangZhou, 363000, Fujian, China (e-mail: jinjinli@mnnu.edu.cn).

Y.H Qian is with the Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006, China (e-mail: jinchengqyh@126.com).

A. Related works

As an important tool of feature selection, rough set theory [8] has attracted considerable attention in classification learning. However, classical rough sets can only deal with symbolic datasets. Continuous and numerical attributes must be discretized before data reduction, which may lead to the loss of classification information [9],[10]. In view of this observation, some researchers focused on the combination of rough sets and fuzzy sets. Fuzzy rough sets (FRS), as proposed by Dubois and Prade [11], provides an effective means for overcoming the puzzle of data discretization and can handle continuous or numerical datasets without preprocessing. Over the past decade, considerable efforts have been devoted to propose various FRS-based feature selection models by following the remarkable work of Dubois and prade [12]-[17]. In these models, fuzzy dependency functions were used to characterize the distinguishing ability of a given feature subset. However, the fuzzy dependency function is always calculated using the nearest sample, which will lead to unstable classification performance for datasets with noise.

At present, most of FRS-based feature selection models are based on a single granulation [21],[22], which may limit their applications and lead to an increase in complexity [23],[24]. Qian et al. [25] generalized Pawlak's rough sets to multigranulation rough sets (MGRS), in which the upper and lower approximations were characterized under multiple equivalence relations. In MGRS, a target concept can be described by a family of combined relations from the perspective of optimism and pessimism. Up to now, MGRS has attracted much attention and a variety of new MGRS models have been reported, including decision-theoretic MGRS [26],[27], fuzzy MGRS [28]-[30], intuitionistic fuzzy MGRS [31], covering based MGRS [32], neighborhood-based MGRS [33], variable precision MGRS [34], and so on.

In addition, considering that FRS-based feature selection models are hard to deal with the information fusion of fuzzy granularity induced by coverings, several proposals have been made to generalize FRS by using the notion of fuzzy covering [35]-[42]. These models and methods can be viewed as bridges between FRS and covering rough sets. Li et al. [35] proposed fuzzy covering-based rough sets, where an implicator norm and a triangular norm were employed to construct approximation operators. Feng et al. [36] investigated the information fusion for multi-fuzzy covering systems by means of a pair of belief and plausibility functions. D'eer et al.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems

[37],[38] discussed the relationship between sixteen covering based fuzzy neighborhood operators. Liu et al. [39] proposed covering based multigranulation fuzzy rough sets. Zhang et al. [40] designed a novel procedure for making decisions with covering-based intuitionistic fuzzy rough sets.

Ma [43],[44] generalized the notion of fuzzy covering to fuzzy β covering by replacing 1 with a parameter β ($0 < \beta \leq 1$). Yang and Hu [45],[46] presented several fuzzy β covering based rough sets. By following the ideas of Ma [43], Zhan et al. [47] put forth a covering based multigranulation fuzzy rough sets model by means of fuzzy β neighborhoods. In recent years, some generalized rough set models related to fuzzy β covering have been proposed and applied to multi-attribute decision making [48],[49].

B. Our work

Most of fuzzy β covering based rough set models contain only a single granulation structure, which limits their application in multi-source datasets and high dimensional datasets [23]. More promising, multigranulation data analysis provides a more flexible and effective means for evaluating the distinguishing ability of a family of coverings. First, multiple granulation structures can characterize the distinguishing information at different granularity levels, which is beneficial to classification learning. Second, multigranulation rough sets are able to transform the issue of information fusion into multigranulation fusion from multiple views and levels. However, a review of the aforementioned studies shows the current research on fuzzy covering based multigranulation rough sets mainly focuses on model generalization [47], rarely involving feature selection and classification learning for real-world data. We illustrate the possible reasons from three aspects.

(1) For most of fuzzy β covering based rough set models [43],[47],[49], the upper and lower approximation operators have no inclusion relation, i.e., the lower approximation is not included in the upper approximation when $\beta \neq 1$. They can not describe the differences between objects accurately, which will lead to unstable performance in classification learning.

(2) As well as FRS model, the fuzzy dependency functions of these models are obtained by the nearest objects. This leads to the sensitivity to noisy data in classification learning. When the condition attributes are with noise (attribute noise) or the decision attribute contains errors (category noise), the training model can not ideally fit the data.

(3) Fuzzy β neighborhood is severed as the basic granularity for fuzzy β covering based rough set model. However, most of these models are usually constructed directly by a single fuzzy covering, which can not reflect the information fusion among multiple fuzzy coverings. In multigranulation applications cases, it is necessary to construct fuzzy β neighborhood by using one or more fuzzy β covering families.

These gaps inspires our investigation on a new fitting model with fuzzy covering based multigranulation rough set models, as well as their applicability to feature selection.

The main contributions of this article are stated as follow. First, a new fuzzy β covering based multigranulation rough sets is introduced. It can overcome the defect of fuzzy β covering based rough set models that there is no inclusion relation between upper and lower approximations, and describe the differences between samples more accurately. Meanwhile, it has a certain resistance to noisy data and provides a more robust manner for feature selection. Second, fuzzy β neighborhood with respect to a family of fuzzy β coverings is used as basic information granules to formulate fuzzy decision of samples, and optimistic and pessimistic fuzzy dependency functions are proposed. Finally, the dimensionality reduction is carried out by view of maintaining the discriminatory power, and a forward algorithm for feature selection is developed by means of the optimistic significance function.

The rest is organized as follows. In Section 2, we briefly review some covering based rough set models and multi-granulation rough set models. In Section 3, a new fuzzy β covering based multigranulation rough set model is introduced. A heuristic feature selection algorithm based on the optimistic dependency function is designed in Section 4. Finally, some experimental tests and conclusions are presented.

II. Basic notions and results

In this section, we review some basic notions related to fuzzy β covering based rough sets and and multigranulation rough sets. Throughout this paper, U denotes a finite and non-empty set, and $\mathcal{F}(U)$ means all fuzzy sets of U.

A. Fuzzy β covering based rough sets

Let $C = \{K_1, K_2, \cdots, K_m\}$ be a nonempty subset of $\mathcal{F}(U)$. We call C a fuzzy β covering of U if $(\bigcup_{i=1}^{m} K_i)(x) \geq \beta$ for each $x \in U$.

The fuzzy β neighborhood of $x \in U$ is formulated as

$$[x]_C^\beta = \bigcap\{K | K \in C, K(x) \geq \beta\}.$$

Definition 1. *[43] Suppose C is a fuzzy β covering of U. For any $X \in \mathcal{F}(U)$, the lower and upper approximations of X w.r.t. C are defined, respectively.*

$$\underline{C}(X)(x) = \bigwedge_{y\in U} \{(1 - [x]_C^\beta(y)) \vee X(y)\},$$

$$\overline{C}(X)(x) = \bigvee_{y\in U} \{[x]_C^\beta(y) \wedge X(y)\},$$

for all $x \in U$, where "$\bigvee$" means "max" and "$\bigwedge$" means "min".

When $\beta = 1$, then the above fuzzy β covering based rough sets becomes a general covering based fuzzy rough sets.

Definition 2. *[49] Suppose C is a fuzzy β covering of U. For any $X \in \mathcal{F}(U)$, the fuzzy lower and upper approximations of X with a variable precision $k \in [0, 1]$ are defined, respectively.*

$$\underline{C}^k(X)(x) = \inf_{X(y)\leq k}\{(1 - [x]_C^\beta(y)) \vee k\} \wedge \inf_{X(y)>k}\{(1 - [x]_C^\beta(y)) \vee X(y)\},$$

$$\overline{C}^k(X)(x) = \inf_{X(y)\geq 1-k}\{[x]_C^\beta(y) \wedge (1 - k)\} \vee \inf_{X(y)<1-k}\{[x]_C^\beta(y) \wedge X(y)\},$$

for all $x \in U$.

B. Multigranulation rough sets

Definition 3. *[25] Let (U, A) be an information system and $A_1, A_2, \cdots, A_n \subseteq A$. The optimistic multigranulation lower and upper approximations of $X \subseteq U$ w.r.t. $A_1, A_2, \cdots, A_n$ are denoted by*

$$\underline{R}^{O}_{\sum_{i=1}^{n} A_i}(X) = \{x \in U | [x]_{A_1} \subseteq X \vee [x]_{A_2} \subseteq X \vee \cdots \vee [x]_{A_n} \subseteq X\},$$

$$\overline{R}^{O}_{\sum_{i=1}^{n} A_i}(X) = \{x \in U | [x]_{A_1} \cap X \neq \varnothing \wedge [x]_{A_2} \cap X \neq \varnothing \wedge \cdots \wedge [x]_{A_n} \cap X \neq \varnothing\}.$$

The pessimistic multigranulation lower and upper approximations of $X \subseteq U$ w.r.t. $A_1, A_2, \cdots, A_n$ are denoted by

$$\underline{R}^{P}_{\sum_{i=1}^{n} A_i}(X) = \{x \in U | [x]_{A_1} \subseteq X \wedge [x]_{A_2} \subseteq X \wedge \cdots \wedge [x]_{A_n} \subseteq X\},$$

$$\overline{R}^{P}_{\sum_{i=1}^{n} A_i}(X) = \{x \in U | [x]_{A_1} \cap X \neq \varnothing \vee [x]_{A_2} \cap X \neq \varnothing \vee \cdots \vee [x]_{A_n} \cap X \neq \varnothing\}.$$

Definition 4. *[47] Let $\Delta = \{C_i, C_2, \cdots, C_n\}$ be a family of fuzzy β coverings of U. For any $X \subseteq U$, denote*

$$\underline{R}^{O,\beta}_{\sum_{i=1}^{n} C_i}(X)(x) = \bigvee_{i=1}^{n} \bigwedge_{y \in U} \{(1 - [x]^{\beta}_{C_i}(y)) \vee X(y)\},$$

$$\overline{R}^{O,\beta}_{\sum_{i=1}^{n} C_i}(X)(x) = \bigwedge_{i=1}^{n} \bigvee_{y \in U} \{[x]^{\beta}_{C_i}(y) \bigwedge X(y)\}.$$

for all $x \in U$.

Then $(\underline{R}^{O,\beta}_{\sum_{i=1}^{n} C_i}(X), \overline{R}^{O,\beta}_{\sum_{i=1}^{n} C_i}(X))$ is called a pair of covering based optimistic multigranulation operators of X.

Similarly, the fuzzy pessimistic multigranulation approximation operators are denoted by

$$\underline{R}^{P,\beta}_{\sum_{i=1}^{n} C_i}(X)(x) = \bigwedge_{i=1}^{n} \bigwedge_{y \in U} \{(1 - [x]^{\beta}_{C_i}(y)) \vee X(y)\},$$

$$\overline{R}^{P,\beta}_{\sum_{i=1}^{n} C_i}(X)(x) = \bigvee_{i=1}^{n} \bigvee_{y \in U} \{[x]^{\beta}_{C_i}(y) \bigwedge X(y)\}.$$

for all $x \in U$.

III. Fuzzy β covering based multigranulation rough sets

In this section, the shortcoming of some existing fuzzy β covering based rough sets are first illustrated. To break through these limitations, a novel fuzzy covering based multigranulation rough set model is then presented.

A. The defects of fuzzy β covering and some of its extended models

The inclusion between upper and lower approximations is an important property of rough set models, which can be used to characterize and approximate a given target concept. Most of rough set models and their variants have the property of inclusion. However, fuzzy β covering based rough sets are obviously different, most of them don't have the property of inclusion when $\beta \neq 1$, such as: Ma et. al. [43], Yang et. al. [45],[46], Zhan et. al. [47], and Jiang et. al. [49].

An example below is employed to illustrate the problem.

Example 1. *Let $U = \{x_1, x_2, x_3\}$ and $C_1 = \{K_{11}, K_{12}, K_{13}\}$, $C_2 = \{K_{21}, K_{22}, K_{23}\}$ be a family of fuzzy sets of U, where*

$$K_{11} = \frac{0.6}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3}, \quad K_{12} = \frac{0.7}{x_1} + \frac{0.2}{x_2} + \frac{0.3}{x_3},$$
$$K_{13} = \frac{0.2}{x_1} + \frac{0.6}{x_2} + \frac{0.7}{x_3}, \quad K_{21} = \frac{0.5}{x_1} + \frac{0.7}{x_2} + \frac{0.3}{x_3},$$
$$K_{22} = \frac{0.3}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3}, \quad K_{23} = \frac{0.7}{x_1} + \frac{0.2}{x_2} + \frac{0.8}{x_3}.$$

Let $\beta = 0.5$ and $X = \frac{0.7}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3}$.

By the definition of fuzzy β neighborhood, we have

$$[x_1]^{\beta}_{C_1} = \frac{0.6}{x_1} + \frac{0.2}{x_2} + \frac{0.3}{x_3}, \quad [x_2]^{\beta}_{C_1} = \frac{0.2}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3},$$
$$[x_3]^{\beta}_{C_1} = \frac{0.2}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3}, \quad [x_1]^{\beta}_{C_2} = \frac{0.5}{x_1} + \frac{0.2}{x_2} + \frac{0.3}{x_3},$$
$$[x_2]^{\beta}_{C_2} = \frac{0.3}{x_1} + \frac{0.5}{x_2} + \frac{0.3}{x_3}, \quad [x_3]^{\beta}_{C_2} = \frac{0.3}{x_1} + \frac{0.2}{x_2} + \frac{0.6}{x_3}.$$

By Definition 1, we compute that

$$\underline{C}_1(X) = \frac{0.7}{x_1} + \frac{0.5}{x_2} + \frac{0.5}{x_3}, \quad \overline{C}_1(X) = \frac{0.6}{x_1} + \frac{0.6}{x_2} + \frac{0.6}{x_3},$$
$$\underline{C}_2(X) = \frac{0.7}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3}, \quad \overline{C}_2(X) = \frac{0.5}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3}.$$

Obviously, we have that
$\underline{C}_1(X) \nsubseteq \overline{C}_1(X)$ and $\underline{C}_2(X) \nsubseteq \overline{C}_2(X)$.

Hence, the upper and lower approximations defined by Ma [43] have no inclusion relation.

Let $k = 0.6$. By Definition 2, we obtain that

$$\underline{C}^k_1(X) = \frac{0.7}{x_1} + \frac{0.6}{x_2} + \frac{0.6}{x_3}, \quad \overline{C}^k_1(X) = \frac{0.6}{x_1} + \frac{0.6}{x_2} + \frac{0.6}{x_3},$$
$$\underline{C}^k_2(X) = \frac{0.7}{x_1} + \frac{0.6}{x_2} + \frac{0.6}{x_3}, \quad \overline{C}^k_2(X) = \frac{0.5}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3}.$$

It is clear that $\underline{C}^k_1(X) \nsubseteq \overline{C}^k_1(X)$ and $\underline{C}^k_2(X) \nsubseteq \overline{C}^k_2(X)$.

Hence, the variable precision upper and lower approximations defined by Jiang et al. [49] have no inclusion relation.

By Definition 4, we compute that

$$\underline{R}^{O,\beta}_{C_1+C_2}(X) = \frac{0.7}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3}, \quad \overline{R}^{O,\beta}_{C_1+C_2}(X) = \frac{0.5}{x_1} + \frac{0.5}{x_2} + \frac{0.6}{x_3},$$
$$\underline{R}^{P,\beta}_{C_1+C_2}(X) = \frac{0.7}{x_1} + \frac{0.5}{x_2} + \frac{0.5}{x_3}, \quad \overline{R}^{P,\beta}_{C_1+C_2}(X) = \frac{0.6}{x_1} + \frac{0.6}{x_2} + \frac{0.6}{x_3}.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems

Obviously, we have $\underline{R}^{O,\beta}_{C_1+C_2}(X) \not\subseteq \overline{R}^{O,\beta}_{C_1+C_2}(X)$ and $\underline{R}^{P,\beta}_{C_1+C_2}(X) \not\subseteq \overline{R}^{P,\beta}_{C_1+C_2}(X)$.

Hence, the multigranulation upper and lower approximations defined by Zhan et al. [47] have no inclusion relation.

It is worth noting that some researchers try to solve this problem by constructing new covering neighborhood. Zhang et al. [50] presented a new type of fuzzy α neighborhood by combing fuzzy β covering and the fuzzy covering neighborhoods proposed by D'eer et al. [37]. However, the computation of multiple neighborhoods will greatly increase the computational complexity.

B. Fuzzy β covering based multigranulation rough sets

In this subsection, a novel multigranulation fuzzy rough set model is presented, and some relative properties are explored.

Definition 5. *Let $\Delta = \{C_1, C_2, \cdots, C_m\}$ be a family of fuzzy β coverings of U. The pair (U, Δ) is called a fuzzy β covering approximation space. Furthermore, (U, Δ, d) is called a fuzzy β covering decision table, where d is a decision attribute.*

Next, we define a pair of multigranulation approximation operators by means of fuzzy β neighborhood. The new proposed model has two advantages. On the one hand, it can guarantee the inclusion relation between the lower and upper approximations, so as to characterize a given objective concept accurately. On the other hand, considering that the weak membership of a sample to the target concept may be caused by noisy data [55], it is reasonable to set the lower approximation as zero when the membership degree is less than $1-\beta$. Thus, it can reduce the influence of noisy data, and better fit the distribution of a given data set.

Definition 6. *Let (U, Δ) be a fuzzy β covering approximation space, and $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_n \subseteq \Delta$. For each $X \in \mathcal{F}(U)$, the fuzzy lower approximation and upper approximation of X are denoted by*

$$\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) = \begin{cases} \bigvee_{i=1}^{n} \bigwedge_{y\in U} \{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee X(y)\}, & X(x) \geq 1-\beta \\ 0, & \textit{otherwise} \end{cases}, \quad (1)$$

$$\overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) = \begin{cases} \bigwedge_{i=1}^{n} \bigvee_{y\in U} \{[x]^{\beta}_{\mathcal{P}_i}(y) \wedge X(y)\}, & X(x) \leq \beta \\ 1, & \textit{otherwise} \end{cases}, \quad (2)$$

where $[x]^{\beta}_{\mathcal{P}_i} = \bigcap_{C\in\mathcal{P}_i} [x]^{\beta}_{C}$, for all $x \in U$.

If $\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X) \neq \overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)$, then X is called an optimistic fuzzy β covering based multigranulation rough sets, otherwise it is optimistic definable.

Remark:

(1) If $\mathcal{P}_i$ is a single point set consisting of a single covering for any $1 \leq i \leq n$, and $\beta = 1$, then the above formulas become as

$$\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) = \bigvee_{i=1}^{n} \bigwedge_{y\in U} \{(1-[x]^{1}_{\mathcal{P}_i}(y)) \vee X(y)\}, \quad (3)$$

$$\overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) = \bigwedge_{i=1}^{n} \bigvee_{y\in U} \{[x]^{1}_{\mathcal{P}_i}(y) \wedge X(y)\}. \quad (4)$$

This means that $(\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X), \overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X))$ can be viewed as a special case of covering based multigranulation rough sets proposed by zhan et al. [47] ($\beta = 1$).

In particular, if $\mathcal{P}_1 = \mathcal{P}_2 = \cdots = \mathcal{P}_n$, then formulas (3) and (4) become as follow:

$$\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) = \bigwedge_{y\in U} \{(1-[x]^{1}_{\mathcal{P}_i}(y)) \vee X(y)\}, \quad (5)$$

$$\overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) = \bigvee_{y\in U} \{[x]^{1}_{\mathcal{P}_i}(y) \wedge X(y)\}. \quad (6)$$

This means that $(\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X), \overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X))$ will degenerate into a fuzzy covering based rough set proposed by Ma [43] ($\beta = 1$).

If a fuzzy relation R_i on U is denoted by $R_i(x, y) = [x]^{1}_{\mathcal{P}_i}(y)$ for any $x, y \in U$, then formula (5) and (6) become

$$\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) = \bigwedge_{y\in U} \{(1-R_i(x,y)) \vee X(y)\}, \quad (7)$$

$$\overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) = \bigvee_{y\in U} \{R_i(x,y) \wedge X(y)\}. \quad (8)$$

This means that $(\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X), \overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X))$ can be regarded as a fuzzy rough set proposed by Dubois and Prade [11].

(2) If $\mathcal{P}_1 = \mathcal{P}_2 = \cdots = \mathcal{P}_n$, then the above formulas (1) and (2) become as

$$\underline{R}^{\beta}_{\mathcal{P}_i}(X)(x) = \begin{cases} \bigwedge_{y\in U} \{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee X(y)\}, & X(x) \geq 1-\beta \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

$$\overline{R}^{\beta}_{\mathcal{P}_i}(X)(x) = \begin{cases} \bigvee_{y\in U} \{[x]^{\beta}_{\mathcal{P}_i}(y) \wedge X(y)\}, & X(x) \leq \beta \\ 1, & \text{otherwise} \end{cases}. \quad (10)$$

This means that $(\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X), \overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X))$ will degenerate into a single-granulation fuzzy covering based rough sets proposed by Huang et al. [4].

We then discuss the inclusion relation between the new approximation operators.

Theorem 1. *Let $\mathcal{P}_1, \mathcal{P}_2 \cdots, \mathcal{P}_n \subseteq \Delta$ and $X \in \mathcal{F}(U)$, then $\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X) \subseteq \overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)$.*

Proof. *We prove Theorem 1 in two cases.*

(1) $\beta < 1-\beta$

For any $x \in U$, if $X(x) < 1-\beta$, we have $\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) = 0$, thus $\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x) \leq \overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(X)(x)$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems

If $X(x) \geq 1-\beta$, then $X(x) > \beta$, we have $\overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) = 1$. Hence, $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) \leq \overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x)$.

(2) $\beta \geq 1-\beta$

For any $x \in U$, if $X(x) < 1-\beta$, then $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) = 0$. Since $\overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) \geq 0$, we have $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) \leq \overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x)$.

If $1-\beta \leq X(x) \leq \beta$, we know that $[x]^{\beta}_{\mathcal{P}_i}(x) \geq \beta$, which implies that $1-[x]^{\beta}_{\mathcal{P}_i}(x) \leq X(x)$.

Thus,

$\bigwedge_{y \in U}\{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee X(y)\} \leq (1-[x]^{\beta}_{\mathcal{P}_i}(x)) \vee X(x) = X(x)$.

It follows that

$\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) = \bigvee_{i=1}^{n} \bigwedge_{y \in U}\{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee X(y)\} \leq X(x)$.

Moreover,

$\bigvee_{y \in U}\{[x]^{\beta}_{\mathcal{P}_i}(y) \wedge X(y)\} \geq [x]^{\beta}_{\mathcal{P}_i}(x) \wedge X(x) = X(x)$,

We obtain that

$\overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) = \bigwedge_{i=1}^{n} \bigvee_{y \in U}\{[x]^{\beta}_{\mathcal{P}_i}(y) \wedge X(y)\} \geq X(x)$.

Hence, $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) \subseteq \overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)$.

If $X(x) > \beta$, $\overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) = 1$, we have $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) \leq \overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x)$.

In summary, we obtain that $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) \subseteq \overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)$.

Proposition 1. *Let $\mathcal{P}_1, \mathcal{P}_2 \cdots, \mathcal{P}_n \subseteq \Delta$ and $X, Y \in \mathcal{F}(U)$. The following properties hold:*

(1L) If $X \subseteq Y$, then $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) \subseteq \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y)$;

(1H) If $X \subseteq Y$, then $\overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) \subseteq \overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y)$;

(2L) $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X \cap Y) = \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) \cap \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y)$;

(2H) $\overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X \cup Y) = \overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) \cup \overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y)$.

Proof. *(1L) If $X \subseteq Y$, then $X(y) \leq Y(y)$ for any $y \in U$. If $X(x) \geq 1-\beta$,*

$$\begin{aligned}\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) &= \bigvee_{i=1}^{n} \bigwedge_{y \in U}\{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee X(y)\} \\ &\leq \bigvee_{i=1}^{n} \bigwedge_{y \in U}\{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee Y(y)\} \\ &= \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y)(x).\end{aligned}$$

If $X(x) < 1-\beta$, $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) = 0 \leq \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y)(x)$,

Hence, $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) \subseteq \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y)$.

(1H) The proof is similar to (1L)

(2L) If $X(x) \geq 1-\beta$ and $Y(x) \geq 1-\beta$, then

$$\begin{aligned}&\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X \cap Y)(x) \\ &= \bigvee_{i=1}^{n} \bigwedge_{y \in U}\{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee (X \cap Y)(y)\} \\ &= \bigvee_{i=1}^{n} \bigwedge_{y \in U}\{((1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee X(y)) \wedge ((1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee Y(y))\} \\ &= \bigvee_{i=1}^{n} \bigwedge_{y \in U}\{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee X(y)\} \wedge \bigvee_{i=1}^{n} \bigwedge_{y \in U}\{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee Y(y)\} \\ &= \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) \wedge \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y)(x) \\ &= (\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) \cap \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y))(x)\end{aligned}$$

If $X(x) < 1-\beta$ or $Y(x) < 1-\beta$, then $(X \cap Y)(x) < 1-\beta$, yield $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X \cap Y)(x) = 0$, and hence,

$$\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X \cap Y)(x) = (\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) \cap \underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(Y))(x) = 0.$$

(2H) The proof is similar to to (2L).

From Definition 6 and formula (9), (10), we can easily obtain the following results.

Proposition 2.

(1) $\underline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) = \bigcup_{i=1}^{n} \underline{R}^{\beta}_{\mathcal{P}_i}(X)$;

(2) $\overline{R}^{O,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X) = \bigcap_{i=1}^{n} \overline{R}^{\beta}_{\mathcal{P}_i}(X)$.

The first item demonstrates that the optimistic multigranulation lower approximation of a target concept is the union of single granulation lower approximations. On the one hand, it shows that the optimistic lower approximation increases monotonically with the size of granularity. The introduction of new fuzzy coverings is beneficial to the accurate description of the target concept. On the other hand, it provides a good way for us to incrementally calculate the optimistic lower approximation. The second item indicates the pessimistic upper approximation of a target concept is the intersection of the single granulation upper approximations.

Definition 7. *Let (U, Δ) be a fuzzy β covering approximation space, $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_n \subseteq \Delta$. For each $X \in \mathcal{F}(U)$, the pessimistic lower and upper approximations of X are denoted by*

$$\underline{R}^{P,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) = \begin{cases} \bigwedge_{i=1}^{n} \bigwedge_{y \in U}\{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee X(y)\}, & X(x) \geq 1-\beta \\ 0, & otherwise \end{cases},$$

$$\overline{R}^{P,\beta}_{\sum_{i=1}^{n} \mathcal{P}_i}(X)(x) = \begin{cases} \bigvee_{i=1}^{n} \bigvee_{y \in U}\{[x]^{\beta}_{\mathcal{P}_i}(y) \wedge X(y)\}, & X(x) \leq \beta \\ 1, & otherwise \end{cases}.$$

IV. Feature selection with covering based multigranulation rough fuzzy sets

In this section, the dimensionality reduction of fuzzy β covering decision tables is explored. In order to allow more samples to enter the positive domain, Wang et al. [55] proposed the concept of fuzzy decision. Next, we extend this idea to fuzzy covering decision table by means of fuzzy β neighborhood.

Definition 8. *Given a fuzzy β covering decision table (U, Δ, d) with $U/R_d = \{D_1, D_2, \cdots, D_r\}$. For any $x \in U$, denote*

$$\tilde{D}_i(x) = \frac{\left|[x]^{\beta}_{\Delta} \cap D_i\right|}{\left|[x]^{\beta}_{\Delta}\right|}, \quad i = 1, 2, \cdots, r, \tag{11}$$

where $\tilde{D}_i$ is a fuzzy set, and $\tilde{D}_i(x)$ means the membership degree of x to D_i. Then $\{\tilde{D}_1, \tilde{D}_2, \cdots, \tilde{D}_r\}$ is called the fuzzy decision of samples induced by d.

Definition 9. *Given a fuzzy β covering decision table (U, Δ, d), and $\{\tilde{D}_1, \tilde{D}_2, \cdots, \tilde{D}_r\}$ is the fuzzy decision of objects induced by d. The optimistic fuzzy multigranulation lower and upper approximations of decision d w.r.t. $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_n$ are defined as*

$$\underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{D}_i)(x) = \begin{cases} \bigvee_{i=1}^{n} \bigwedge_{y\in U} \{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee \tilde{D}_i(y)\}, & \tilde{D}_i(x) \geq 1-\beta \\ 0, & otherwise \end{cases}$$

$$\overline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{D}_i)(x) = \begin{cases} \bigwedge_{i=1}^{n} \bigvee_{y\in U} \{[x]^{\beta}_{\mathcal{P}_i}(y) \wedge \tilde{D}_i(y)\}, & \tilde{D}_i(x) \leq \beta \\ 1, & otherwise \end{cases}.$$

Accordingly, the pessimistic fuzzy multigranulation approximations w.r.t $\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_n$ are defined as

$$\underline{R}^{P,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{D}_i)(x) = \begin{cases} \bigwedge_{i=1}^{n} \bigwedge_{y\in U} \{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee \tilde{D}_i(y)\}, & \tilde{D}_i(x) \geq 1-\beta \\ 0, & otherwise \end{cases}$$

$$\overline{R}^{P,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{D}_i)(x) = \begin{cases} \bigvee_{i=1}^{n} \bigvee_{y\in U} \{[x]^{\beta}_{\mathcal{P}_i}(y) \wedge \tilde{D}_i(y)\}, & \tilde{D}_i(x) \leq \beta \\ 1, & otherwise \end{cases}.$$

Then the optimistic fuzzy positive domain and dependency function of d are formulated as

$$POS^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{d}) = \bigcup_{i=1}^{r} \underline{R}^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{D}_i), \tag{12}$$

$$\partial^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{d}) = \frac{\sum_{x\in U} POS^{O,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{d})(x)}{|U|}. \tag{13}$$

Similarly, the pessimistic fuzzy positive domain and dependency function are denoted by

$$POS^{P,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{d}) = \bigcup_{i=1}^{r} \underline{R}^{P,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{D}_i), \tag{14}$$

$$\partial^{P,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{d}) = \frac{\sum_{x\in U} POS^{P,\beta}_{\sum_{i=1}^{n}\mathcal{P}_i}(\tilde{d})(x)}{|U|}. \tag{15}$$

Remark: In formulas (12) and (13), the fuzzy multigranulation low approximation is used to construct the fuzzy dependency function to evaluate the distinguishing ability of a family of coverings. Although it doesn't use the property of inclusion between upper and lower approximations directly, the approximation operators in Definition 9 can fit a given data set well. It overcomes the shortcoming that traditional fuzzy rough set can't guarantee that the maximal membership of an object to its own category [55], and provides an effective means of preventing the misclassification of training samples. Moreover, the property of inclusion is an important property for a rough set model, which can be used to characterize and approximate a given target concept.

Example 2. *Let (U, Δ, d) be a fuzzy β covering decision table, where $U = \{x_1, x_2, x_3\}$, $\Delta = \{C_1, C_2, C_3, C_4\}$ be a family of fuzzy β coverings of U, $C_i = \{K_{i1}, K_{i2}, K_{i3}\}$, $i = 1, 2, 3, 4$, $\mathcal{P}_1 = \{C_1, C_2\}$, $\mathcal{P}_2 = \{C_3, C_4\}$, $U/R_d = \{D_1, D_2\}$, $D_1 = \{x_1, x_2\}$, $D_2 = \{x_3\}$ and*

$$K_{11} = \frac{0.8}{x_1} + \frac{0.6}{x_2} + \frac{0}{x_3}, \quad K_{12} = \frac{0.7}{x_1} + \frac{1}{x_2} + \frac{0.2}{x_3},$$
$$K_{13} = \frac{0.1}{x_1} + \frac{0.8}{x_2} + \frac{0.9}{x_3}, \quad K_{21} = \frac{0.8}{x_1} + \frac{0.7}{x_2} + \frac{0.3}{x_3},$$
$$K_{22} = \frac{0.7}{x_1} + \frac{1}{x_2} + \frac{0.3}{x_3}, \quad K_{23} = \frac{0.3}{x_1} + \frac{0.8}{x_2} + \frac{1}{x_3},$$
$$K_{31} = \frac{0.9}{x_1} + \frac{1}{x_2} + \frac{0.3}{x_3}, \quad K_{32} = \frac{1}{x_1} + \frac{1}{x_2} + \frac{0.3}{x_3},$$
$$K_{33} = \frac{0.3}{x_1} + \frac{0.8}{x_2} + \frac{1}{x_3}, \quad K_{41} = \frac{0.8}{x_1} + \frac{0.3}{x_2} + \frac{0}{x_3},$$
$$K_{42} = \frac{0}{x_1} + \frac{1}{x_2} + \frac{0.8}{x_3}, \quad K_{43} = \frac{0}{x_1} + \frac{0.8}{x_2} + \frac{1}{x_3}.$$

Let $\beta = 0.5$, we would like to find $\partial^{O,\beta}_{\mathcal{P}_1+\mathcal{P}_2}(\tilde{d})$ and $\partial^{P,\beta}_{\mathcal{P}_1+\mathcal{P}_2}(\tilde{d})$.
By the definition of fuzzy β neighborhood, we compute that

$$[x_1]^{\beta}_{C_1} = \frac{0.7}{x_1} + \frac{0.6}{x_2} + \frac{0}{x_3}, \quad [x_2]^{\beta}_{C_1} = \frac{0.1}{x_1} + \frac{0.6}{x_2} + \frac{0}{x_3},$$
$$[x_3]^{\beta}_{C_1} = \frac{0.1}{x_1} + \frac{0.8}{x_2} + \frac{0.9}{x_3}, \quad [x_1]^{\beta}_{C_2} = \frac{0.7}{x_1} + \frac{0.7}{x_2} + \frac{0.3}{x_3},$$
$$[x_2]^{\beta}_{C_2} = \frac{0.3}{x_1} + \frac{0.7}{x_2} + \frac{0.3}{x_3}, \quad [x_3]^{\beta}_{C_2} = \frac{0.3}{x_1} + \frac{0.8}{x_2} + \frac{1}{x_3},$$
$$[x_1]^{\beta}_{C_3} = \frac{0.9}{x_1} + \frac{1}{x_2} + \frac{0.3}{x_3}, \quad [x_2]^{\beta}_{C_3} = \frac{0.3}{x_1} + \frac{0.8}{x_2} + \frac{0.3}{x_3},$$
$$[x_3]^{\beta}_{C_3} = \frac{0.3}{x_1} + \frac{0.8}{x_2} + \frac{1}{x_3}, \quad [x_1]^{\beta}_{C_4} = \frac{0.8}{x_1} + \frac{0.3}{x_2} + \frac{0}{x_3},$$
$$[x_2]^{\beta}_{C_4} = \frac{0}{x_1} + \frac{0.8}{x_2} + \frac{0.8}{x_3}, \quad [x_3]^{\beta}_{C_4} = \frac{0}{x_1} + \frac{0.8}{x_2} + \frac{0.8}{x_3}.$$

Subsequently,

$$[x_1]^{\beta}_{\Delta} = \frac{0.7}{x_1} + \frac{0.3}{x_2} + \frac{0}{x_3}, \quad [x_2]^{\beta}_{\Delta} = \frac{0}{x_1} + \frac{0.6}{x_2} + \frac{0}{x_3},$$
$$[x_3]^{\beta}_{\Delta} = \frac{0}{x_1} + \frac{0.8}{x_2} + \frac{0.8}{x_3}.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems

By Definition 8, we have

$$\tilde{D}_1 = \frac{1}{x_1} + \frac{1}{x_2} + \frac{0.5}{x_3}, \qquad D_2 = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.5}{x_3}.$$

We then compute that

$$[x_1]^{\beta}_{\mathcal{P}_1} = \frac{0.7}{x_1} + \frac{0.6}{x_2} + \frac{0}{x_3}, \quad [x_2]^{\beta}_{\mathcal{P}_1} = \frac{0.1}{x_1} + \frac{0.6}{x_2} + \frac{0}{x_3},$$
$$[x_3]^{\beta}_{\mathcal{P}_1} = \frac{0.1}{x_1} + \frac{0.8}{x_2} + \frac{0.9}{x_3}, \quad [x_1]^{\beta}_{\mathcal{P}_2} = \frac{0.8}{x_1} + \frac{0.3}{x_2} + \frac{0}{x_3},$$
$$[x_2]^{\beta}_{\mathcal{P}_2} = \frac{0}{x_1} + \frac{0.8}{x_2} + \frac{0.3}{x_3}, \quad [x_3]^{\beta}_{\mathcal{P}_2} = \frac{0}{x_1} + \frac{0.8}{x_2} + \frac{0.8}{x_3}.$$

It is calculated by Definition 9 that

$$\underline{R}^{\beta}_{\mathcal{P}_1}\left(\tilde{D}_1\right) = \frac{1}{x_1} + \frac{1}{x_2} + \frac{0.5}{x_3}, \ \underline{R}^{\beta}_{\mathcal{P}_1}\left(\tilde{D}_2\right) = \frac{0.3}{x_1} + \frac{0.4}{x_2} + \frac{0.2}{x_3},$$
$$\underline{R}^{\beta}_{\mathcal{P}_2}\left(\tilde{D}_1\right) = \frac{1}{x_1} + \frac{0.7}{x_2} + \frac{0.5}{x_3}, \ \underline{R}^{\beta}_{\mathcal{P}_2}\left(\tilde{D}_2\right) = \frac{0.2}{x_1} + \frac{0.2}{x_2} + \frac{0.2}{x_3}.$$

It follows that

$$\underline{R}^{O,\beta}_{\mathcal{P}_1+P_2}\left(\tilde{D}_1\right) = \frac{1}{x_1} + \frac{1}{x_2} + \frac{0.5}{x_3}, \ \underline{R}^{O,\beta}_{\mathcal{P}_1+P_2}\left(\tilde{D}_2\right) = \frac{0.3}{x_1} + \frac{0.4}{x_2} + \frac{0.2}{x_3},$$
$$\underline{R}^{P,\beta}_{\mathcal{P}_1+P_2}\left(\tilde{D}_1\right) = \frac{1}{x_1} + \frac{0.7}{x_2} + \frac{0.5}{x_3}, \ \underline{R}^{P,\beta}_{\mathcal{P}_1+P_2}\left(\tilde{D}_2\right) = \frac{0.2}{x_1} + \frac{0.2}{x_2} + \frac{0.2}{x_3},$$

From formula (13), we can determine that

$$\partial^{O,\beta}_{\mathcal{P}_1+\mathcal{P}_2}(\tilde{d}) \approx 0.83, \qquad \partial^{P,\beta}_{\mathcal{P}_1+\mathcal{P}_2}(\tilde{d}) \approx 0.73.$$

Theorem 2. *Let (U, Δ, d) be a fuzzy β covering decision table, $\mathcal{F}(\Delta)$ be the power set of Δ, and $\mathcal{P} \subseteq \mathcal{Q} \subseteq \mathcal{F}(\Delta)$, then*

(1) $\partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d}) \leq \partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{Q}} \mathcal{P}_i}(\tilde{d})$;

(2) $\partial^{P,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{Q}} \mathcal{P}_i}(\tilde{d}) \geq \partial^{P,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d})$.

Proof. *(1) If $\tilde{D}_i(x) \geq 1-\beta$, since $\mathcal{P} \subseteq \mathcal{Q}$, we have*

$$\begin{aligned}
\underline{R}^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{Q}} \mathcal{P}_i}(\tilde{D}_i)(x) &= \bigvee_{\mathcal{P}_i \in \mathcal{Q}} \bigwedge_{y \in U} \{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee \tilde{D}_i(y)\} \\
&= \{\bigvee_{\mathcal{P}_i \in \mathcal{P}} \bigwedge_{y \in U} \{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee \tilde{D}_i(y)\}\} \\
&\quad \vee \{\bigvee_{\mathcal{P}_i \in \mathcal{Q}-\mathcal{P}} \bigwedge_{y \in U} \{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee \tilde{D}_i(y)\}\} \\
&\geq \bigvee_{\mathcal{P}_i \in \mathcal{P}} \bigwedge_{y \in U} \{(1-[x]^{\beta}_{\mathcal{P}_i}(y)) \vee \tilde{D}_i(y)\} \\
&= \underline{R}^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{D}_i)(x).
\end{aligned}$$

If $\tilde{D}_i(x) < 1-\beta$, we obtain $\underline{R}^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{Q}} \mathcal{P}_i}(\tilde{D}_i)(x) = \underline{R}^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{D}_i)(x) = 0$.

Thus, $\underline{R}^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{D}_i) \subseteq \underline{R}^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{Q}} \mathcal{P}_i}(\tilde{D}_i)$.

By formula (12), we have $POS^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d}) \subseteq POS^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{Q}} \mathcal{P}_i}(\tilde{d})$.

Hence, $\partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d}) \leq \partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{Q}} \mathcal{P}_i}(\tilde{d})$.

(2) The proof is similar to (1).

The first item reveals that the optimistic fuzzy dependent function monotonically increases with the size of the feature subset. It ensures that adding new features to the existing feature set will not reduce the dependency function, which provides a basis for us to design a forward algorithm for feature selection. If the dependent function is no longer increased, the search for candidate features will stop. The second item indicates that pessimistic fuzzy dependent function decreases monotonously with the numbers of features.

Theorem 3. *Let (U, Δ, d) be a fuzzy β covering decision table, $\beta_1 < \beta_2$ and $\mathcal{P} \subseteq \mathcal{F}(\Delta)$, then*

(1) $\partial^{O,\beta_1}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d}) \geq \partial^{O,\beta_2}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d})$;

(2) $\partial^{P,\beta_1}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d}) \geq \partial^{P,\beta_2}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d})$.

Proof. *(1) As $\beta_1 < \beta_2$, we have $[x]^{\beta_1}_{\mathcal{P}} \subseteq [x]^{\beta_2}_{\mathcal{P}}$. If $\tilde{D}_i(x) \geq 1-\beta$, for any $y \in U$, we can obtain that $1-[x]^{\beta_2}_{\mathcal{P}}(y) \leq 1-[x]^{\beta_1}_{\mathcal{P}}(y)$. From the definition of optimistic lower approximation, we have* $\underline{R}^{O,\beta_2}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{D}_i) \subseteq \underline{R}^{O,\beta_1}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{D}_i)$.

It follows that $POS^{O,\beta_2}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d}) \subseteq POS^{O,\beta_1}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d})$.

Hence, $\partial^{O,\beta_2}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d}) \leq \partial^{O,\beta_1}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d})$.

(2) The proof is similar to (1).

Theorem 3 demonstrates that the optimistic and pessimistic fuzzy dependent functions monotonically change with the value of parameter β.

Next, we discuss the knowledge reduction of fuzzy β covering decision tables.

Definition 10. *Let (U, Δ, d) be a fuzzy β covering decision table, and $\mathcal{P} \subseteq \mathcal{F}(\Delta)$. For any $\mathcal{P}' \in \mathcal{P}$, if* $\partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}-\{\mathcal{P}'\}} \mathcal{P}_i}(\tilde{d}) = \partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d})$*, we say $\mathcal{P}'$ is optimistic redundant in $\mathcal{P}$. Otherwise, we say $\mathcal{P}'$ is optimistic indispensable. If any $\mathcal{P}'$ in $\mathcal{P}$ is optimistic indispensable, we call $\mathcal{P}$ is optimistic independent.*

Redundant fuzzy coverings can not improve the discriminatory power of covering families, and even interfere with classification learning. Therefore, they must be reduced before the training of classifiers.

Definition 11. *Let (U, Δ, d) be a fuzzy β covering decision table, and $\mathcal{P} \subseteq \mathcal{Q} \subseteq \mathcal{F}(\Delta)$. We say $\mathcal{P}$ is an optimistic reduct of $\mathcal{Q}$, if it satisfies*

1) $\partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d}) = \partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{Q}} \mathcal{P}_i}(\tilde{d})$;

2) $\partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}-\{\mathcal{P}'\}} \mathcal{P}_i}(\tilde{d}) < \partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d})$, $\forall \mathcal{P}' \in \mathcal{P}$.

By Definition 11, we know that the reducts refer to the minimum fuzzy β covering families, which has the same discriminatory power as the whole covering family.

Definition 12. *Let (U, Δ, d) be a fuzzy β covering decision table, $\mathcal{P} \subseteq \mathcal{F}(\Delta)$, and $\mathcal{P}' \in \mathcal{F}(\Delta) - \mathcal{P}$. The optimistic significance degree of $\mathcal{P}'$ with respect to $\mathcal{P}$ is defined as*

$$SD^{O,\beta}(\mathcal{P}', \mathcal{P}, d) = \partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P} \cup \{\mathcal{P}'\}} \mathcal{P}_i}(\tilde{d}) - \partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i \in \mathcal{P}} \mathcal{P}_i}(\tilde{d}).$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems

Based on the aforementioned discussion, we formulate a heuristic feature selection algorithm by means of the optimistic significance degree.

Algorithm 1 A heuristic algorithm for feature selection with fuzzy β covering based multigranulation rough sets (FBCMG)

Input: $S=(U,\Delta,d)$ and $\mathcal{Q}\subseteq\mathcal{F}(\Delta)$
Output: A reduct $\mathcal{P}$
1: Initialize $\mathcal{P}=\varnothing$;
2: **for** $\mathcal{P}'\in\mathcal{Q}-\mathcal{P}$ **do**
3: Calculate $\underline{R}^{\beta}_{\mathcal{P}'}(\tilde{D}_j)$, for $1\leq j\leq r$
4: Calculate $\underline{R}^{O,\beta}_{\sum\limits_{\mathcal{P}_i\in\mathcal{P}\cup\{\mathcal{P}'\}}\mathcal{P}_i}(\tilde{D}_j)$, for $1\leq j\leq r$,
5: Obtain $\partial^{O,\beta}_{\sum\limits_{\mathcal{P}_i\in\mathcal{P}\cup\{\mathcal{P}'\}}\mathcal{P}_i}(\tilde{d})$ by formula (13);
6: Obtain the optimistic significance degree $SD^{O,\beta}(\mathcal{P}',\mathcal{P},d)$ by Definition 12;
7: **end for**
8: Find $\mathcal{P}'$ maximizing $SD^{O,\beta}(\mathcal{P}',\mathcal{P},d)$;
9: **if** $SD^{O,\beta}(\mathcal{P}',\mathcal{P},d)>0$ **then**
10: $\mathcal{P}\leftarrow\mathcal{P}\cup\{\mathcal{P}'\}$;
11: $\mathcal{Q}\leftarrow\mathcal{Q}-\{\mathcal{P}'\}$;
12: Goto Step 2;
13: **else**
14: Return the reduct $\mathcal{P}$;
15: **end if**

Next, the time complexity of the new algorithm is discussed. In Step 3, $\underline{R}^{\beta}_{\mathcal{P}'}(\tilde{D}_j)$ is calculated for each $\tilde{D}_j$ with the time complexity $O((\sum\limits_{P\in\mathcal{P}'}|P|)|U|^2)$. In Step 4, $\underline{R}^{O,\beta}_{\sum\limits_{\mathcal{P}_i\in\mathcal{P}\cup\{\mathcal{P}'\}}\mathcal{P}_i}(\tilde{D}_j)$ can be computed by using an incremental strategy. $\underline{R}^{O,\beta}_{\sum\limits_{\mathcal{P}_i\in\mathcal{P}}\mathcal{P}_i}(\tilde{D}_j)$ is stored in the previous cycle. By Proposition 2, we just need to find the union between it and $\underline{R}^{\beta}_{\mathcal{P}'}(\tilde{D}_j)$. So the time complexity of Step 4 is $O(|U|^2)$. In Step 5, the optimistic multigranulation dependency function can be obtained within $O(|U|^2)$. In Steps 9–15, the optimistic significance of each fuzzy covering family in $\mathcal{Q}$ can be measured with the complexity $O((\sum\limits_{\mathcal{P}'\in\mathcal{Q}}\sum\limits_{P\in\mathcal{P}'}|P|)|U|^2)$. Thus, the overall computational complexity of the algorithm is $O((\sum\limits_{\mathcal{P}'\in\mathcal{Q}}\sum\limits_{P\in\mathcal{P}'}|P|)|U|^2)$.

V. Numerical experiment

The main goal of feature selection includes two aspects, one is to select the optimal feature subset in a more robust case; the other is to obtain higher classification performance on the reduced data. In this section, three groups of simulation experiments are employed to verify the effectiveness and feasibility of the proposed model. First, we evaluated the robustness of five fuzzy β covering based rough set models against noisy data. Second, the dependency functions were constructed from different granularity levels by selecting different neighborhood parameters, so as to obtain optimal feature subsets. Finally, the classification performance of the proposed model was evaluated by comparing with some state-of-the-art feature selection algorithms.

TABLE I: DESCRIPTION OF DATA SETS

No	Data sets	Instances	Features	Classes
1	Wdbc	569	30	2
2	Ionos	351	30	2
3	Sonar	208	60	2
4	Cleve	296	13	2
5	WBC	683	9	2
6	Appendicitis	106	7	2
7	German	1000	24	2
8	Breast	277	9	2
9	Vote	435	16	2
10	Wine	178	13	3
11	DLBCLTumor	77	7129	2
12	DLBCLOutcome	58	7129	2
13	ColonTumor	62	2000	2
14	AMLALL-Total	72	7129	2
15	DLBCLStanford	47	4026	2
16	NervousSystem	60	7229	2

These experiments are conducted on 16 real-word data sets, including ten UCI data sets [51] (Wdbc, Ionos, Sonar, Cleve, WBC, Appendicitis, German, Breast, Vote and Wine) and six high-dimensional gene data sets [52] (DLBCLTumor, DLBCLOutcome, ColonTumor, AMLALL-Total, DLBCLStanford and NervousSystem). The detailed information are displayed in Table I. These UCI datasets can be obtained from UCI machine learning repository (https://archive.ics.uci.edu/ml/index.php), and the gene data sets can be downloaded from the Elvira biomedical data set repository (http://leo.ugr.es/elvira/DBCRepository/).

All experiments are implemented by MATLAB 2016b simulation environment, which is installed in a mobile workstation with windows 10 operating system, i7 CPU and 16GB RAM. Three classical machine learning classifiers including K-nearest neighbors (KNN, K=10), Classification and regression trees (CART) and Naive bayes classifier (NBC), are employed to demonstrate the classification performance. These classifiers are provided by the machine learning toolbox of MATLAB. The comparison experiment of reduced data is realized by using 10-fold cross validation, in which each data set is randomly divided into ten parts of the same size. Then, each part is used in rotation as the test dataset, and the other as the train dataset. The 10-fold cross validation experiment is repeated ten times, and the mean and variance of accuracy are taken as evaluation index.

A. Robustness comparison

In some security-sensitive practical applications, the success of rough set models depend on their resistance to noisy data. For a robust rough set model, we wish that the positive region or dependent function should not be greatly affected by the data disturbance, that is, when some values of attributes change slightly, the positive region should not change dramatically. In [53], D'eer et al. presented an error index by means of positive region, using which what extent a given rough set model is sensitive to noise can be evaluated. If one membership of $n\%$ fuzzy coverings changes to random value in the range, the noise level is called as n. By following the idea of D'eer et

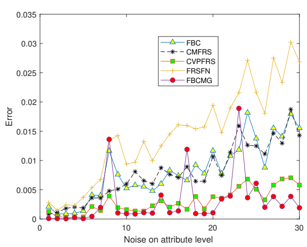
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems

al., we give an evaluation index to examine the robustness of fuzzy β covering based rough sets.

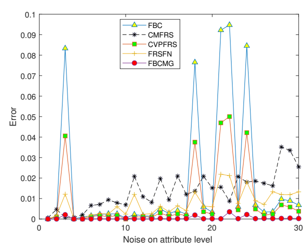
Let $\beta \in (0, 1]$, the change degree of positive region when $n\%$ noise is imposed on a given data set is defined as

$$error_{\beta}^{n} = \frac{\sum_{x \in U} \left| POS_{\beta}(x) - POS_{\beta}^{n}(x) \right|}{|U|},$$

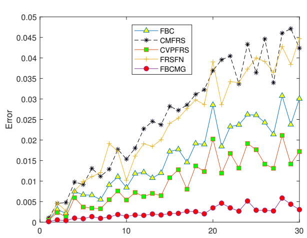
where $POS_{\beta}^{n}(x)$ and $POS_{\beta}(x)$ mean the altered positive domain and the raw positive domain under a given β, respectively. Obviously, the change degree can reflect the anti-noise ability of a given model. The smaller the change degree is, the stronger the robustness of the model becomes.



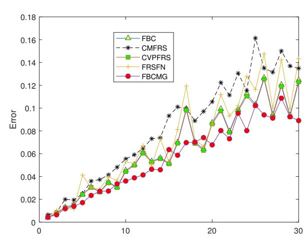
(a) Wdbc dataset



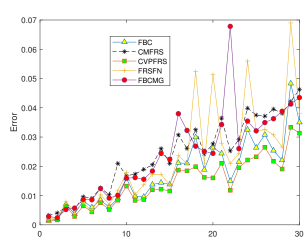
(b) Ionos dataset



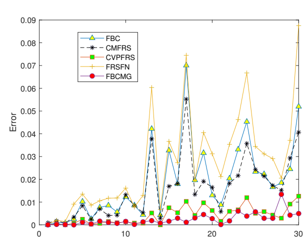
(c) Sonar dataset



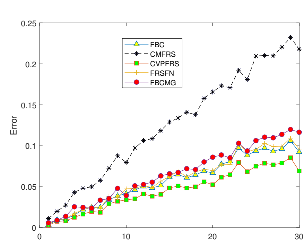
(d) Cleve dataset



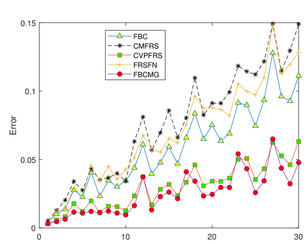
(e) WBC dataset



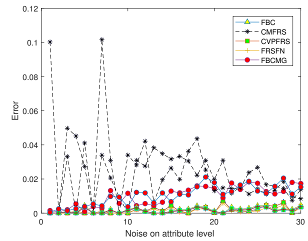
(f) Appendicitis dataset



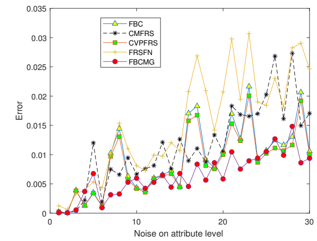
(g) German dataset



(h) Breast dataset



(i) Vote dataset



(j) Wine dataset

Fig. 1: Robustness test on UCI data sets

TABLE II: The average change degree of positive region on UCI datasets (%)

Data sets	FBC	CMFRS	CVPFRS	FRSFN	FBCMG
Wdbc	0.821	0.796	0.309	1.448	**0.308**
Ionos	1.721	1.358	0.904	0.758	**0.045**
Sonar	1.570	2.601	1.006	2.396	**0.231**
Cleve	6.501	8.278	6.410	7.232	**5.614**
WBC	1.759	2.381	**1.431**	2.251	2.383
Appendicitis	1.867	1.564	0.413	2.707	**0.240**
German	5.782	12.707	**4.611**	5.986	6.579
Breast	5.795	7.585	3.008	6.939	**2.529**
Vote	0.191	2.214	0.150	**0.148**	1.047
Wine	0.920	1.129	0.856	1.523	**0.640**

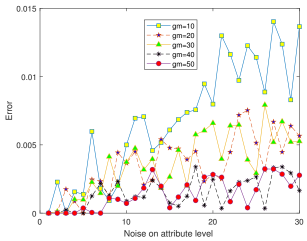
We compare the proposed model (FBCMG) with some state-of-the-art fuzzy covering based rough set models, which have been discussed in Section III (A). These models include fuzzy β covering based rough sets (FBC) [43], covering based multigranulation fuzzy rough sets (CMFRS) [47], covering based variable precision fuzzy rough sets (CVPFRS) [49], and fuzzy rough sets with fuzzy neighborhood (FRSFN) [50]. These models are all constructed on the basis of fuzzy β neighborhood. In the experiment, we set the parameter $\beta = 0.75$. In multigranulation rough sets, one granularity may have one or more attributes. In order to simulate the multigranulation scene, each granularity is composed of two attributes. If the number of condition attributes cannot be divisible by 2, the last attribute constitutes a granularity.

The noise level n is set to vary from 1 to 30 with a step of 1. Fig. 1 shows the variation curve of error index on different data sets. All values are the mean of ten repeated runs. The X-aixs shows the noise level n, and the error is indicated in Y-axis. Accordingly, the less the curve varies is, the better the robustness becomes. It is clear that FBCMG outperforms the other three models for most cases. In particular, FBCMG performs remarkably better for Ionos, Sonar and Appendicitis datasets.

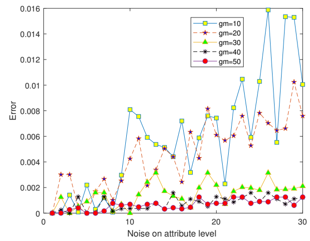
Table II presents the average change degree of positive region, where the data in bold means the minimum of change degree. Out of the total ten datasets, FBCMG achieves the lowest change degree in 7 cases, CVPFRS and FRSFN attain it for 2 cases and 1 case, respectively. It is clear that the FBCMG is more robust against noise than the other three models.

One of the advantages of multigranulation data analysis lies in its ability to characterize knowledge at different granularity levels. To reflect the influence of knowledge at different

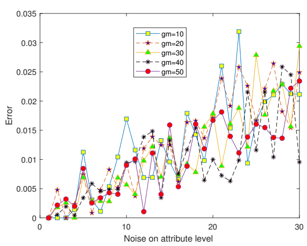
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems



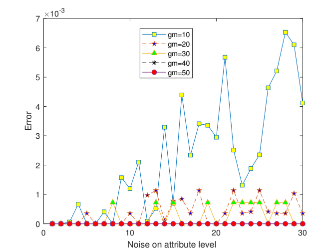
(a) DLBCLTumor dataset



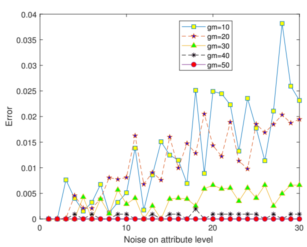
(b) DLBCLOutcome dataset



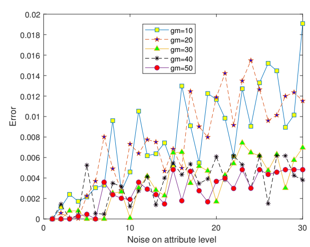
(c) ColonTumor dataset



(d) AMLALL-Total dataset



(e) DLBCLStanford dataset



(f) NervousSystem dataset

Fig. 2: Robustness test on high-dimension data sets with different granularity

granularity on the robustness, we construct granularity with different sizes for high-dimensional datasets. Suppose that the number of attributes in each granularity is gm, we set gm to a value between 10 to 50 in a step of 10. Fig. 2 shows the varied curve of positive region under different granularity. We can choose the appropriate granularity level based on these figures. Table III lists the average change degree of positive region with four different models. For the total six datasets, FBCMG obtains the lowest change degree in 5 cases. This shows that our model is more robust for high dimensional data sets which are more sensitive to attribute noise.

TABLE III: The average change degree of positive region on high-dimension datasets (%)

Data sets	FBC	CMFRS	CVPFRS	FRSFN	FBCMG ($gm=50$)
DLBCLTumor	0.612	1.073	**0.026**	0.902	0.165
DLBCLOutcome	0.095	1.039	0.086	0.609	**0.075**
ColonTumor	0.892	1.263	0.865	1.852	**0.832**
AMLALL-Total	0.001	0.001	**0.00**	0.019	**0.00**
DLBCL-Stanford	0.182	**0.00**	0.173	0.501	**0.00**
NervousSystem	0.583	1.435	0.577	0.783	**0.229**

B. Effect of covering neighborhood parameter β

We know the size of fuzzy β neighborhood is closely related to the size of β. Different β means that the data can be observed from different granularity levels. In order to discuss the influence of different parameters on classification performance, the notion of reduction rate is given.

Definition 13. *Let $\beta \in (0,1]$, the reduction rate of FBCMG with respect to β is denoted by*

$$rate^{\beta} = 1 - \frac{|C^{\beta}|}{|C|},$$

where $|C|$ and $|C^{\beta}|$ are the number of all features and the number of selected features under a given β, respectively. Obviously, a higher reduction rate means fewer features are selected.

It is well known that the size of selected feature subsets and classification accuracy are two important aspects to evaluate the reduction quality of feature selection models. Therefore, it is necessary to select appropriate value of β to obtain the optimal reduction performance. In the experiment, to simulate the multigranulation scene, each granularity is composed of two attributes. The feature subset and reduction rate of FBCMG model under different parameters are first obtained. The classification accuracy on the reduced data is then computed by 10-fold cross validation under KNN, CART and NBC.

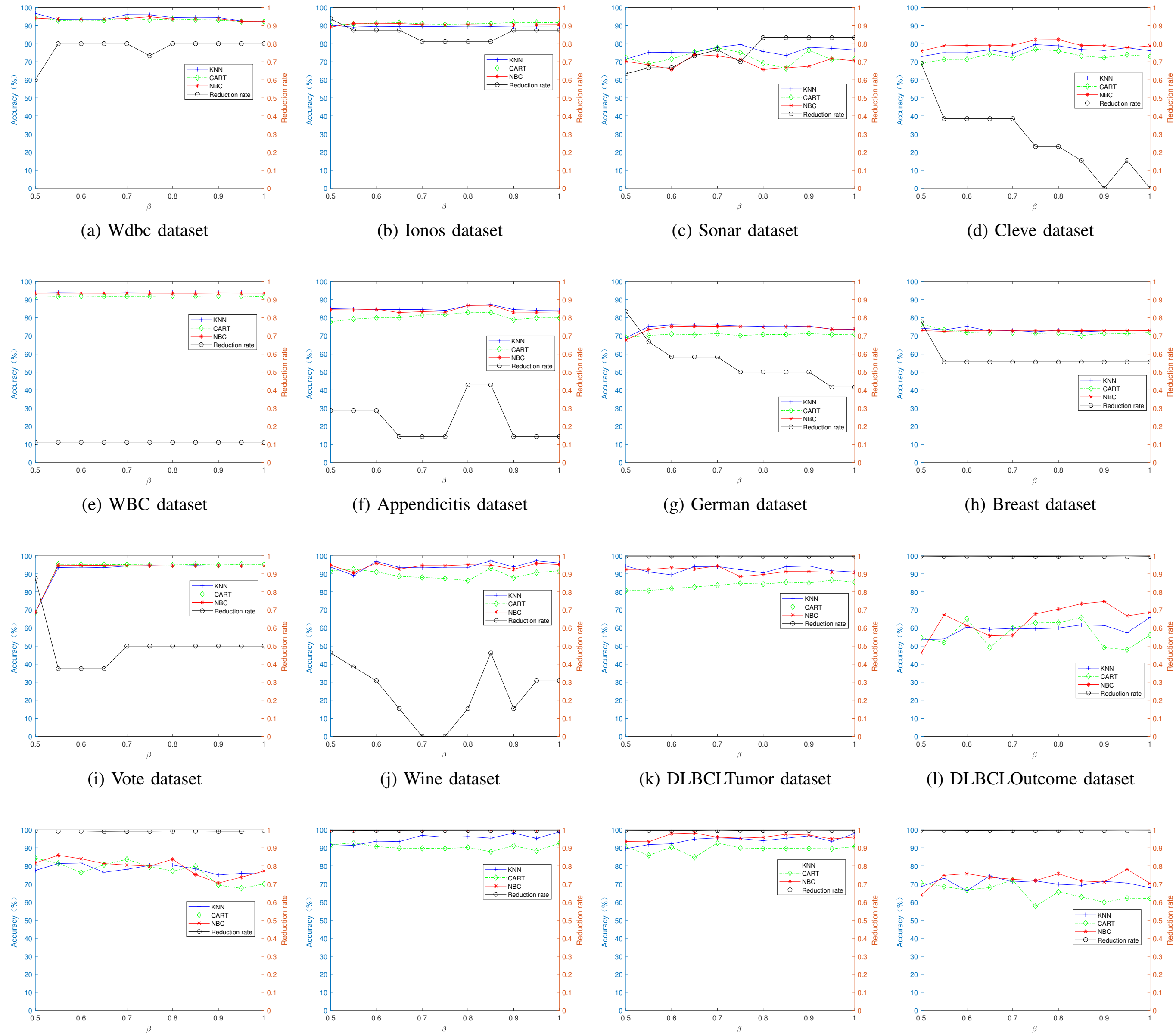
TABLE IV: The optimal value of β with different classifiers

Data sets	KNN	CART	NBC
Wdbc	0.5	0.5	0.75
Ionos	0.5	0.9	0.6
Sonar	0.75	0.7	0.65
Cleve	0.75	0.75	0.80
WBC	0.95	0.8	0.90
Appendicitis	0.85	0.8	0.85
German	0.6	0.7	0.65
Breast	0.6	0.5	0.95
Vote	0.85	0.55	0.65
Wine	0.95	0.85	0.6
DLBCLTumor	0.9	0.95	0.7
DLBCLOutcome	1	0.85	0.9
ColonTumor	0.6	0.5	0.55
AMLALL-Total	1	0.55	1
DLBCL-Stanford	1	0.7	0.65
NervousSystem	0.75	0.7	0.95

For FBCMG, β is set to a value between 0.5 and 1 with a step of 0.05. The changes of classification accuracy and reduction rate are shown in Fig. 3, where, the x-axis shows different values of β, and the left and right y-axis, indicate the classification accuracy and reduction rate, respectively. We can find an optimal parameter for each dataset in Fig. 3. Table IV lists the optimal β of different datasets when FBCMG attains the highest classification accuracy.

For example, Fig. 3 (a) shows the classification accuracy of Wdbc dataset under different classifiers, and the value of β can be set as 0.5, 0.5, 0.75 under KNN, CART and NBC, respectively. From Fig. 3 (b), the classification accuracy varies with the increase of β, and it can achieve the highest accuracy for Ionos dataset when β is set to 0.5, 0.9, 0.6 under KNN,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems



(a) Wdbc dataset (b) Ionos dataset (c) Sonar dataset (d) Cleve dataset

(e) WBC dataset (f) Appendicitis dataset (g) German dataset (h) Breast dataset

(i) Vote dataset (j) Wine dataset (k) DLBCLTumor dataset (l) DLBCLOutcome dataset

(m) ColonTumor dataset (n) AMLALL-Total dataset (o) DLBCLStanford dataset (p) NervousSystem dataset

Fig. 3: The classification accuracy varying with β

CART and NBC, respectively. For FBCMG, all the results in the following tables of Section V(C) were obtained with the optimal parameters shown in Table IV.

C. Classification results of different models

In this subsection, we compare the FBCMG model with several popular feature selection methods, including (1) multi-granulation entropy based feature selection (MGEFS) [54]; (2) Covering based optimistic multigranulation fuzzy rough sets (CMFRS) [47]; (3) Fitting model with fuzzy rough sets (FMFRS) [55].

Firstly, the size of feature subsets obtained by the four models on each dataset is compared. Table V shows the number of selected features by different models. The average dimensionality reduction of FBCMG reaches 62.3% for the first ten UCI datasets, and 99.6% for the other six high dimensional gene datasets. On the whole, the average number of attributes selected by FBCMG under different classifiers are fewer than those of CMFRS and FMFRS model, and slightly inferior to MGEFS. Table VI lists the optimal feature subset of FBCMG with three classifiers.

In the following, we compare the classification performance of FBCMG with MGEFS, CMFRS, and FMFRS. Following the models and experiments designed in [47],[54],[55], the average and variance of classification accuracy on KNN, CART and NBC classifiers are shown in Table VII to Table IX, respectively. From table V, it can be seen that the features selected by different models are quite different. From

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems

TABLE V: The average sizes of selected feature subsets

Data sets	Raw data	MGEFS	CMFRS	FMFRS	FBCMG		
					KNN	CART	NBC
Wdbc	30	2	14	4	12	12	8
Ionos	30	30	10	1	2	4	4
Sonar	60	37	23	8	18	14	16
Cleve	13	3	4	8	10	10	10
WBC	9	4	5	5	8	8	8
Appendicitis	7	6	3	2	4	4	4
German	24	5	4	17	10	10	10
Breast	9	8	5	5	4	2	4
Vote	16	9	4	9	8	10	10
Wine	13	3	7	8	9	7	9
DLBCLTumor	7129	8	75	26	20	18	14
DLBCLOutcome	7129	8	65	34	28	32	34
ColonTumor	2000	7	51	24	14	10	14
AMLALL-Total	7129	21	69	30	28	18	28
DLBCL-Stanford	4026	3	48	15	14	8	10
NervousSystem	7229	7	66	20	20	22	32
Average	2178.3	10.1	28.3	13.5	13.1	11.8	13.4

Table VII to Table IX, we further analyze the differences in classification accuracy. It is clear that the classification performance of FBCMG outperforms that of other three models on most of datasets. In total 48 instances, FBCMG achieves the highest accuracy 30 times, FMFRS 11 times, CMFRS 4 times and MGEFS 4 times. Furthermore, the average accuracy of FBCMG is the highest under different classifiers. Compared with raw data, the average accuracy has been greatly improved, by 6.4% for KNN, 5.4% for CART , and 6.9% for NBC.

Although MGEFS select few features for some datasets, the accuracy of MGEFS is far inferior to other models. Some important features may be deleted in the reduction process, which leads to the decrease of classification accuracy.

As we know, no model can always better than the others for different learning tasks. In general, our model can reduce redundant features and perform better than other three models for most of datasets.

D. Statistical Analysis

In this subsection, two hypothesis tests i.e., Friedman test [56] and Bonferoni-Dunn test [57] are used to systematically explore the classification performance of different algorithms. For Friedman test, the statistics are expressed as:

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2} \text{ and } \chi_F^2 = \frac{12N}{k(k+1)}\left(\sum_{i=1}^{k} R_i^2 - \frac{k(k+1)^2}{4}\right),$$

where k and N are the number of algorithms and datasets, respectively; R_i denotes the mean rank of a given algorithm in all datasets; and F_F follows a Fisher distribution with $F(k-1, k-1(N-1))$ freedom degrees.

TABLE X: Statistical test of four models with three classifiers

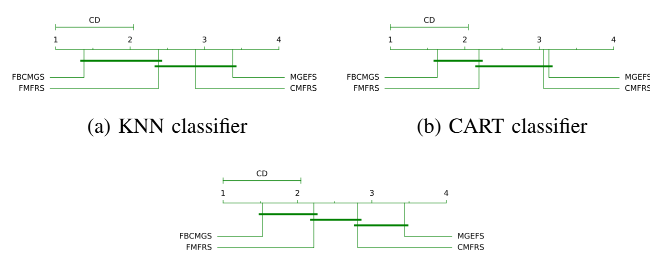
Classifiers	Average ranking				χ_F^2	F_F
	MGEFS	CMFRS	FMFRS	FBCMG		
KNN	3.38	2.88	2.38	1.38	21	11.67
CART	3.13	3.06	2.19	1.63	15.07	6.87
NBC	3.44	2.81	2.22	1.53	19.14	9.95

Following the steps of statistical test in [56], we can obtain the mean ranking of each algorithm by averaging the ranking in all datasets. For a given datasets, the best ranking with the highest accuracy is set as 1, and the second ranking is set as 2, and so on. Then we can compute the values of χ_F^2 and F_F. Table X lists the mean ranking of four algorithms and the values of χ_F^2 and F_F under three classifiers. At level $\alpha = 0.1$, the critical value of $F(3, 45)$ is 2.21. From Tale X, we can see that the values of F_F under KNN, CART and NBC classifiers are all larger than $F(3, 45)$. According to Friedman test, the null hypotheses are rejected, and the classification performance of four algorithms are clearly different under KNN, CART and NBC classifiers, respectively.

Therefore, the corresponding post-hoc tests, i.e., Bonferroni-Dunn test is employed to explore the differences of the four algorithms. In Bonferroni-Dunn test, the critical value is described as

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

If the distance of mean ranking between two algorithms is larger than the critical value CD_α, then there is significant differences in the performance of the two algorithms. To intuitively exhibit these differences, a special graph with the critical value is used to connect these algorithms. If there is a link between the two algorithms, it shows that there is no significant difference from each other.



(a) KNN classifier (b) CART classifier

(c) NBC classifier

Fig. 4: Accuracy comparison with four models on three classifiers

Fig. 4 indicates the comparison of FBCMG with other three algorithms under different classifiers, where the line at the top indicates the critical value, and the mean ranking of each algorithm is shown in the axis. In [57], the critical value can be obtained that $CD_{0.1} = 1.05 \ \ (k = 4, N = 16)$.

As seen in Fig. 4 (a), FBCMG performs clearly better than MGEFS and CMFRS at level $\alpha = 0.1$. At the same time, there is no obvious difference between MGEFS, CMFRS and FMFRS. We conclude that the classification performance of FBCMG is better than the others. In terms of CART classifier in Fig. 4 (b), FBCMG attains the best mean ranking, and is significantly better than MGEFS and CMFRS since the difference of mean ranking is larger than the critical value 1.05. Similarly, as shown in Fig. 4 (c) on NBC classifier, FBCMG outperforms MGEFS and CMFRS, and is similar to FMFRS.

To sum up, FBCMG surpasses the other three algorithms as the whole under the results of the Friedman test and Bonferroni-Dunn test.

TABLE VI: The optimal feature subset of FBCMG with three classifiers

Data sets	KNN (K=10)	CART	NBC
Wdbc	27 28 21 22 25 26 9 10 23 24 15 16	27 28 21 22 25 26 9 10 23 24 15 16	27 28 1 2 7 8 21 22
Ionos	25 26	17 18 25 26	25 26 29 30
Sonar	35 36 9 10 15 16 25 26 49 50 11 12 31 32 13 14 5 6	35 36 15 16 45 46 11 12 31 32 49 50 37 38	35 36 11 12 15 16 49 50 31 32 37 38 53 54 33 34
Cleve	3 4 9 10 1 2 11 12 7 8	3 4 9 10 1 2 11 12 7 8	3 4 9 10 7 8 11 12 1 2
WBC	1 2 5 6 3 4 7 8	1 2 5 6 3 4 7 8	1 2 5 6 3 4 7 8
Appendicitis	5 6 3 4	5 6 3 4	5 6 3 4
German	9 10 1 2 3 4 17 18 5 6	3 4 9 10 1 2 5 6 17 18	9 10 3 4 5 6 17 18 1 2
Breast	5 6 3 4	5 6	5 6 7 8
Vote	3 4 11 12 15 16 5 6	3 4 11 12 15 16 5 6 9 10	3 4 11 12 15 16 5 6 9 10
Wine	1 2 13 7 8 11 12 9 10	1 2 13 9 10 7 8	7 8 9 10 1 2 13 11 12
DLBCLTumor	1091 1092 6179 6180 5997 5998 4027 4028 6757 6758 1173 1174 5533 5534 4077 4078 4929 4930	5881 5882 5997 5998 4027 4028 2481 2482 4023 4024 6757 6758 4077 4078 5533 5534 6377 6378	1091 1092 4027 4028 6179 6180 6757 6758 4023 4024 5997 5998 1173 1174
DLBCLOutcome	1323 1324 1395 1396 2139 2140 5463 5464 39 40 4269 4270 1737 1738 405 406 49 50 6705 6706 2257 2258 5847 5848 6013 6014 3271 3272	6641 6642 1395 1396 5993 5994 5847 5848 39 40 49 50 4269 4270 6825 6826 2029 2030 2257 2258 749 750 3271 3272 989 990 1053 1054 4111 4112 2011 2012	5463 5464 749 750 1395 1396 5847 5848 4269 4270 39 40 6825 6826 2029 2030 49 50 2257 2258 6631 6632 989 990 2011 2012 1053 1054 4111 4112 6303 6304 5809 5810
ColonTumor	43 44 1771 1772 493 494 1247 1248 1671 1672 1221 1222 1339 1340	1891 1892 1771 1772 493 494 125 126 467 468	43 44 1891 1892 1771 1772 493 494 639 640 1671 1672 1221 1222
NAMLALL-Total	757 758 4049 4050 1881 1882 5171 5172 2013 2014 2401 2402 683 684 51 52 1779 1780 1809 1810 6277 6278 1669 1670 4679 4680 1685 1686	757 758 1881 1882 4049 4050 1685 1686 1779 1780 4679 4680 2401 2402 6277 6278 2013 2014	757 758 4049 4050 1881 1882 5171 5172 2013 2014 2401 2402 683 684 51 52 1779 1780 1809 1810 6277 6278 1669 1670 4679 4680 1685 1686
DLBCL-Stanford	1275 1276 3861 3862 1317 1318 1281 1282 75 76 3085 3086 3313 3314	1275 1276 2463 2464 2439 2440 1205 1206	1281 1282 1317 1318 3861 3862 75 76 1205 1206
NervousSystem	4587 4588 1477 1478 5977 5978 6679 6680 1053 1054 4605 4606 1807 1808 3041 3042 4649 4650 2757 2758	4587 4588 1477 1478 5977 5978 1053 1054 4605 4606 3041 3042 6679 6680 6135 6136 4649 4650 5847 5848 4941 4942	1477 1478 3185 3186 5977 5978 1047 1048 1053 1054 653 654 3041 3042 6077 6078 5847 5848 3239 3240 2495 2496 4605 4606 599 600 3651 3652 4889 4890 6411 6412

TABLE VII: Comparison of classification accuracies of reduced data with KNN (K=10) (%)

Data Sets	Raw data	MGEFS	CMFRS	FMFRS	FBCMG
Wdbc	96.81±2.42	92.72±3.23	94.30±3.32	94.21±2.53	**96.82±2.16**
Ionos	89.18±1.13	89.20±1.36	89.18±1.25	89.84±4.22	**90.20±4.73**
Sonar	74.29±9.51	69.82±9.64	78.43±8.80	79.47±8.76	**79.49±8.29**
Cleve	78.74±7.73	57.00±8.68	73.59±7.23	79.25±7.07	**79.51±8.01**
WBC	96.69±2.05	96.58±2.42	95.59±2.16	96.62±2.31	**97.16±1.97**
Appendicitis	86.75±8.63	**87.45±9.15**	86.28±8.60	84.77±8.59	87.32±8.07
German	71.18±3.51	70.02±2.73	72.55±3.45	71.26±2.86	**76.06±3.25**
Breast	72.75±6.11	**75.43±5.80**	70.65±6.65	73.13±6.89	75.21±5.74
Vote	92.39±3.89	92.42±3.75	94.14±3.01	91.85±3.78	**94.53±3.24**
Wine	96.56±3.99	90.70±6.87	90.79±6.58	97.01±3.78	**97.12±3.99**
DLBCLTumor	85.14±8.70	87.09±12.45	93.32±9.06	**96.71±5.66**	94.32±8.66
DLBCLOutcome	54.48±21.22	56.60±13.87	56.13±18.42	64.61±18.68	**65.83±16.69**
ColonTumor	72.67±11.10	63.98±15.05	80.57±15.85	79.76±13.44	**81.64±16.48**
AMLALL-Total	79.73±11.78	60.85±12.34	**99.43±1.51**	94.76±7.71	98.89±3.50
DLBCLStanford	75.12±18.79	91.35±13.62	94.47±10.13	**98.05±5.70**	97.92±6.21
NervousSystem	60.79±19.90	58.30±21.15	66.96±17.69	**81.11±14.63**	74.51±16.25
Average	80.20±8.78	77.47±8.88	83.52±7.73	85.78±7.29	**86.66±7.33**

TABLE VIII: Comparison of classification accuracies of reduced data with CART (%)

Data Sets	Raw data	MGEFS	CMFRS	FMFRS	FBCMG
Wdbc	93.12±3.57	90.68±3.71	91.49±3.30	93.23±3.40	**94.47±2.99**
Ionos	88.89±4.63	88.86±4.59	88.29±4.64	88.69±4.24	**91.74±4.49**
Sonar	71.52±8.41	73.61±9.54	71.26±10.20	75.09±9.60	**77.73±10.29**
Cleve	75.79±6.69	53.86±8.46	64.36±8.30	**79.36±6.38**	76.96±6.77
WBC	95.04±2.37	95.37±2.33	94.41±2.63	**95.65±2.37**	95.05±2.29
Appendicitis	82.64±9.81	82.84±10.33	82.61±9.89	76.51±11.39	**83.02±9.12**
German	70.63±4.94	65.05±4.37	67.84±4.36	70.59±4.25	**71.28±3.85**
Breast	66.29±8.54	68.05±7.59	66.82±6.88	73.76±7.20	**76.53±5.79**
Vote	95.01±3.55	**95.84±3.59**	95.36±3.08	93.75±3.56	95.28±3.05
Wine	90.16±6.20	86.97±8.28	91.04±6.31	91.13±6.38	**91.92±6.15**
DLBCLTumor	85.18±11.41	84.89±13.31	87.38±12.09	**90.65±9.54**	86.58±11.19
DLBCLOutcome	53.21±18.83	60.88±18.10	**70.05±22.93**	66.30±19.10	65.70±18.98
ColonTumor	74.62±15.97	67.50±15.38	74.79±16.88	75.19±15.47	**84.31±13.20**
AMLALL-Total	83.61±11.16	63.47±14.87	87.03±8.98	**92.83±8.63**	92.77±8.44
DLBCLStanford	78.07±18.17	88.53±15.20	86.35±14.80	84.95±15.26	**92.73±11.65**
NervousSystem	57.21±16.83	71.45±17.00	61.20±19.77	72.22±20.46	**72.25±19.12**
Average	78.81±9.44	77.37±9.79	80.02±9.69	82.49±9.20	**84.27±8.59**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2021.3093202, IEEE Transactions on Fuzzy Systems

TABLE IX: Comparison of classification accuracies of reduced data with NBC (%)

Data Sets	Raw data	MGEFS	CMFRS	FMFRS	FBCMG
Wdbc	93.46±3.11	90.84±3.57	90.91±3.27	93.49±2.97	**94.92±3.18**
Ionos	85.45±6.81	85.57±6.33	86.15±6.03	89.07±1.37	**91.29±4.20**
Sonar	67.58±10.00	**74.60±10.07**	68.22±10.67	74.05±8.44	74.00±9.14
Cleve	81.09±6.59	60.20±8.67	70.03±7.43	80.84±6.64	**82.22±7.85**
WBC	96.15±2.15	95.86±2.54	95.43±2.22	96.05±2.30	**96.55±2.00**
Appendicitis	85.30±10.67	87.10±9.35	**88.29±8.57**	84.88±9.15	86.89±8.90
German	72.75±3.79	71.15±2.93	71.96±3.56	72.82±4.23	**75.36±4.19**
Breast	72.76±8.14	72.63±6.79	72.33±7.63	**73.05±8.19**	72.95±6.87
Vote	93.98±3.49	93.68±3.91	94.55±3.13	93.52±3.51	**94.75±3.19**
Wine	96.84±3.81	90.16±7.23	92.71±6.30	**96.92±4.29**	95.78±5.01
DLBCLTumor	80.66±13.43	87.30±12.28	87.62±12.38	91.94±9.40	**94.31±8.168**
DLBCLOutcome	44.17±19.13	61.53±17.90	63.00±17.67	58.30±22.94	**74.67±17.94**
ColonTumor	57.48±17.17	57.38±18.57	79.33±16.80	**88.36±14.00**	85.93±14.50
AMLALL-Total	98.48±4.25	62.45±10.71	99.88±0.40	**100.00±0.00**	**100.00±0.00**
DLBCLStanford	94.32±10.94	92.75±13.36	96.07±8.85	97.78±7.01	**98.22±5.64**
NervousSystem	64.24±20.16	60.34±19.77	**82.00±14.31**	73.16±17.41	78.10±15.89
Average	80.29±8.98	77.72±9.62	83.66±8.08	85.26±7.62	**87.25±7.29**

VI. CONCLUSION

Fuzzy β covering is an important means that enables researchers to analyze data in a more general manner. MGRS has been proved to be a powerful tool for characterizing the uncertainty information of knowledge at different granularity levels. In this paper, a novel fitting model for feature selection has been introduced by combining fuzzy β covering and MGRS. On the one hand, this new model ensures the inclusion relationship between the upper and lower approximations, so as to better fit the real data. On the other hand, it can reduce the influence of noisy data and select features in a more robust way. The optimistic fuzzy dependency function is employed as the importance evaluation index of a given fuzzy covering, and the data reduction of fuzzy decision tables is carried out from the perspective of maintaining the discrimination power of the whole fuzzy covering family. Three groups of experiments are used to verify the validity and feasibility of the proposed model. The experimental results show that: (1) Compared with four state-of-the-art fuzzy β covering based rough set models, FBCMG is more robust against data noise. (2) FBCMG performs better than some classical feature selection algorithms in terms of classification accuracy and the size of selected feature subset.

However, some problems need to be further investigated. For example, for FBCMG, the lower approximation is used to construct the dependency function, so as to evaluate the classification ability of feature subsets. How to make full use of the classification information implied in the upper approximation will be an interesting work. Moreover, the proposed model can not fully balance the classification accuracy and the number of selected features for all datasets. We will investigate a more optimal balance algorithm in the future work.

REFERENCES

[1] X. Zhang, C.L. Mei, D.G. Chen, Y.Y. Yang, and J.H. Li, “Active incremental feature selection using a fuzzy rough set-based information entropy, ” IEEE Trans. Fuzzy Syst., vol. 28, no. 5, pp. 901–915, 2020.

[2] C.Z. Wang, Y. Huang, M.W. Shao, Q.H. Hu, and D.G. Chen, “Feature selection based on neighborhood self-information, ” IEEE Trans. Cybern., vol. 50, no. 9, pp. 4031-4042, 2020.

[3] L. Sun, L.Y. Wang, W.P. Ding, Y.H. Qian, J.C. Xu, “Feature selection using fuzzy neighborhood entropy-based uncertainty measures for fuzzy neighborhood multigranulation rough sets, ” IEEE Trans. Fuzzy Syst., Early access, DOI: 10.1109/TFUZZ.2020.2989098.

[4] Z.H. Huang, J.J. Li, “A fitting model for attribute reduction with fuzzy β covering, ” Fuzzy Sets Syst., Early access, DOI: 10.1016/j.fss.2020.07.010.

[5] C.Z. Wang, Q.H. Hu, X.Z. Wang, D.G. Chen, Y.H. Qian, Z. Dong, “Feature selection based on neighborhood discrimination index, ” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 7, pp. 2986–2999, 2018.

[6] L.H. Kong, W.H. Qu, J.D. Yu, H. Zuo, G.H. Chen, F. Xiong, S.R. Pan, S.Y. Lin, M.K. Qiu, “Distributed feature selection for big data using fuzzy rough sets, ” IEEE Trans. Fuzzy Syst., vol. 28, no. 5, pp. 846-857, 2020.

[7] A.H. Tan, W.Z. Wu, Y.H. Qian, J.Y. Liang, J.K. Chen, J.J. Li, “Intuitionistic fuzzy rough set-based granular structures and attribute subset selection, ” IEEE Trans. Fuzzy Syst., vol. 27, no. 3, pp. 527–539, 2019.

[8] Z. Pawlak, “Rough sets, ” Int. J. Comput. Inf. Sci., vol. 11, no. 5, pp. 341-356, 1982.

[9] J.K. Chen, J.S. Mi, and Y.J. Lin, “A graph approach for fuzzy-rough feature selection, ” Fuzzy Sets Syst., vol. 391, pp. 96–116, 2020.

[10] Q.H. Hu, L. Zhang, S. An, D. Zhang, D.R. Yu, “On robust fuzzy rough set models, ” IEEE Trans. Fuzzy Syst., vol. 20, no. 4, pp. 636–651, 2012.

[11] D. Dubois, H. Prade, “Rough fuzzy sets and fuzzy rough sets, ” Int. J. Gen. Syst., vol. 17, pp. 191–209, 1990.

[12] Q.H. Hu, D.R. Yu, and Z.X. Xie, “Information-preserving hybrid data reduction based on fuzzy-rough techniques, ” Pattern Recog. Lett., vol. 27, pp. 414-423, 2006.

[13] R. Jensen, Q. Shen, “New approaches to fuzzy-rough feature selection, ” IEEE Trans. Fuzzy Syst., vol. 17, no. 4, pp. 824–838, 2009.

[14] D.G. Chen, Y.Y. Yang, “Attribute reduction for heterogeneous data based on the combination of classical and fuzzy rough set models, ” IEEE Trans. Fuzzy Syst., vol. 22, no. 5, pp. 1325–1334, 2014.

[15] Y.Y. Yang, D.G. Chen, H. Wang, X.Z. Wang, “Incremental perspective for feature selection based on fuzzy rough sets, ” IEEE Trans. Fuzzy Syst., vol. 26, no. 3, pp. 1257–1273, 2018.

[16] N.M. Parthaláin, R. Jensen, R. Diao, “Fuzzy-rough set bireducts for data reduction, ” IEEE Trans. Fuzzy Syst., vol. 28, no. 8, pp. 1840–1850, 2020.

[17] C.Z. Wang, Y. Wang, M.W. Shao, Y.H. Qian, and D.G. Chen, “Fuzzy rough attribute reduction for categorical data, ” IEEE Trans. Fuzzy Syst., vol. 28, no. 5, pp. 818–830, 2020.

[18] S. Zhao, E.C.C. Tsang, D. Chen, “The model of fuzzy variable precision rough sets, ” IEEE Trans. Fuzzy Syst., vol. 17, no. 2, pp. 451–467, 2009.

[19] J.M. Fernández Salido, S. Murakami, “Rough set analysis of a general type of fuzzy data using transitive aggregations of fuzzy similarity relations, ” Fuzzy Sets Syst., vol. 139, pp. 635–660, 2003.

[20] C. Cornelis, N. Verbiest, R. Jensen, “Ordered weighted average based fuzzy rough sets, ” in: Proceedings of the 5th International Conference on Rough Sets and Knowledge Technology, RSKT2010, pp.78–85, 2010.

[21] D.G. Chen, Q.H. Hu and Y.P. Yang, "Parameterized attribute reduction with gaussian kernel based fuzzy rough sets, " Inf. Sci., vol. 181, no. 23, pp. 5169-5179, 2011.
[22] H. Zhao, P. Wang, Q.H. Hu, "Fuzzy rough set based feature selection for large-scale hierarchical classification, " IEEE Trans. Fuzzy Syst., vol. 27, no. 10, pp. 1891-1903, 2019.
[23] Y.H. Qian, H. Zhang, Y.L. Sang, J.Y. Liang, "Multigranulation decision-theoretical rough sets, " Int. J. Approx. Reason., vol. 55, pp. 225–237, 2014.
[24] A.H. Tan, W.Z. Wu, J.J. Li, T.J. Li, "Reduction foundation with multigranulation rough sets using discernibility, "Artif. Intell. Rev., vol. 53, pp. 2425–2452, 2020.
[25] Y.H. Qian, J. Liang, Y.Y. Yao, C. Dang, "MGRS: a multi-granulation rough set, " Inf. Sci., vol. 180, pp. 949–970, 2010.
[26] Y.T. Guo, E.C.C. Tsang, W.H. Xu, D.G. Chen, "Adaptive weighted generalized multi-granulation interval-valued decision-theoretic rough sets, " Knowl.-Based Syst., vo. 187, 104804, 2020.
[27] Y.H. Qian, X.Y. Liang, G.P. Lin, Qian. Guo, J.Y. Liang, "Local multi-granulation decision-theoretic rough sets, " Int. J. Approx. Reason., vol. 82, pp. 119-137, 2017.
[28] B.Z. Sun, W. Ma, Y.H. Qian, "Multigranulation fuzzy rough set over two universes and its application todecision making, " Knowl-Based Syst., vol. 123, pp. 61–74, 2017.
[29] W.H. Xu, Q.R. Wang, X.T. Zhang, "Multi-granulation fuzzy rough sets in a fuzzy tolerance approximation space, " Int. J. Fuzzy Syst., vol. 13, no. 4, pp. 246–259, 2011.
[30] W.H. Xu, Q.R. Wang, S.Q. Luo, "Multi-granulation fuzzy rough setse, " Int. J. Fuzzy Syst., vol. 26, no. 3, pp. 1323–1340, 2014.
[31] B. Huang, C.X. Guo, Y.L. Zhang, H.X. Li, X.Z. Zhou, "Intuitionistic fuzzy multigranulation rough sets, " Inf. Sci., vol. 277, pp. 299–320, 2014.
[32] Q.Z. Kong, W.H. Xu, "The comparative study of covering rough sets and multi-granulation rough sets, " Soft Comput., vol. 23, pp. 3237–3251, 2019.
[33] G.P. Lin, Y.H. Qian, T.J. Li, "NMGS: neighborhood-based multigranulation rough sets, " Int. J. Approx. Reason., vol. 53, no. 7, pp. 1080–1093, 2012.
[34] X.L. Ma, J.M. Zhan, B.Z. Sun J.C.R. Alcantud, "Novel classes of coverings based multigranulation fuzzy rough sets and corresponding applications to multiple attribute group decision-making, " Artif. Intell. Rev., vol. 53, pp. 6197–6256, 2020.
[35] T.J. Li, Y. Leung, W.X. Zhang, "Generalized fuzzy rough approximation operators based on fuzzy coverings, " Int. J. Approx. Reason., vol. 48, pp. 836–856, 2008.
[36] T. Feng, S.P. Zhang, J.S. Mi, "The reduction and fusion of fuzzy covering systems based on the evidence theory, " Int. J. Approx. Reason., vol. 53, pp. 87–103, 2012.
[37] L. D'eer, C. Cornelis, L. Godo, "Fuzzy neighborhood operators based on fuzzy coverings, " Fuzzy Sets Syst., vol. 312, pp. 17–35, 2017.
[38] L. D'eer, C. Cornelis, "A comprehensive study of fuzzy covering-based rough set models: definitions, properties and interrelationships, " Fuzzy Sets Syst., vol. 336, pp. 1–26, 2018.
[39] C.H. Liu, W. Pedryc, "Covering-based multi-granulation fuzzy rough sets, " Int. J. Fuzzy Syst., vol. 30, pp. 303–318, 2016.
[40] L. Zhang, J.M. Zhan, Z.H. Xu, J.R. Alcantud, "Covering-based general multigranulation intuitionistic fuzzy rough sets and corresponding applications to multi-attribute group decision-making, " Inf. Sci., vol. 494, pp. 114–140, 2019.
[41] B. Yang, B.Q. Hu, "Matrix representations and interdependency on L-fuzzy covering-based approximation operators, " Int. J. Approx. Reason., vol. 96, pp. 57–77, 2018.
[42] X.D. Yue, Y.F. Chen, D.Q. Miao, "Fuzzy neighborhood covering for three-way classification, " Inf. Sci., vol. 507, pp. 795–808, 2020.
[43] L.W. Ma, "Two fuzzy covering rough set models and their generalizations over fuzzy lattices, " Fuzzy Sets Syst., vol. 294, pp. 1–17, 2016.
[44] L.W. Ma, "On some types of neighborhood-related covering rough sets, " Int. J. Approx. Reason., vol. 53, pp. 901–911, 2012.
[45] B. Yang, B.Q. Hu, "On some types of fuzzy covering-based rough sets, " Fuzzy Sets Syst., vol. 312, pp. 36–65, 2017.
[46] B. Yang, B.Q. Hu, "A fuzzy covering-based rough set model and its generalization over fuzzy lattice, " Inf. Sci., vol. 367–368, pp. 463–486, 2016.
[47] J.M. Zhan, X.H. Zhang, Y.Y. Yao, "Covering based multigranulation fuzzy rough sets and corresponding applications, " Artif. Intell. Rev., vol. 53, pp. 1093–1126, 2020.
[48] J.M. Zhan, B.Z. Sun, J.C.R. Alcantud, "Covering based multigranulation (I, T)-fuzzy rough set models and applications in multi-attribute group decision-making, " Inf. Sci., vol. 476, pp. 290–318, 2019.
[49] H.B. Jiang, J.M. Zhan, D.G. Chen, "Covering-based variable precision (I, T)-fuzzy rough sets with applications to multiattribute decision-making, " IEEE Trans. Fuzzy Syst., vol. 27, pp. 1558–1572, 2019.
[50] K. Zhang, J.M. Zhan, W.Z. Wu, "On multi-criteria decision-making method based on a fuzzy rough set model with fuzzy α-neighborhoods, " IEEE Trans. Fuzzy Syst., Early access, DOI 10.1109/TFUZZ.2020.3001670.
[51] C.L. Blake and C.J. Merz, "UCI Repository of Machine Learning Databases, " 1998. [Online]. Available: http://archive.ics.uci.edu/ml/index.php.
[52] Kent Ridge Bio-medical Dataset, http://leo.ugr.es/elvira/DBCRepository/.
[53] L. D'eer, N. Verbiest, C. Cornelis, L. Godo, "A comprehensive study of implicator-conjunctor-based and noise-tolerant fuzzy rough sets:definitions, properties and robustness analysis, " Fuzzy Sets Syst., vol. 275, pp. 1–38, 2015.
[54] K. Zeng, K. She, and X. Z. Niu, "Multi-granulation entropy and its applications, " Entropy, vol. 15, no. 6, pp. 2288–2302, 2013.
[55] C.Z. Wang, Y.L. Qi, M.W. Shao, Q.H. Hu, D.G. Chen, Y.H. Qian, Y.J. Lin, "A fitting model for feature selection with fuzzy rough sets, " IEEE Trans. Fuzzy Syst., vol. 25, pp. 741–753, 2017.
[56] 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.
[57] O.J. Dunn, "Multiple comparisons among means, " J. Am. Stat. Assoc., vol. 56, pp. 52-64, 1961.

Zhehuang Huang received his B.S. degree from Minnan Normal University, Zhangzhou, China, the M.S. degree from Fuzhou University, Fuzhou, China, and the Ph.D. degree from Xiamen University, Xiamen, China, in 2002, 2005, 2015, respectively. He is currently an Associate Professor at the School of Mathematics Sciences with Huaqiao University, China. His research interests are focused on fuzzy sets, rough sets, natural language processing and machine learning. He has authored or coauthored more than 30 journal and conference papers in the areas of machine learning, natural language processing, and rough set theory.

Jinjin Li received his B.S. degree from Guangxi University, Guangxi, China in 1988, and the Ph.D. degree in school of Mathematics and System Sciences from Shandong University, Shandong, China, in 2000. Currently, he is a professor in the School of Mathematics and Statistics, Minnan Normal University. He has published more than 70 articles in international journals and book chapters. His research interests are in the area of data mining, rough sets and topologies.

Yuhua Qian received the M.S. and Ph.D. degrees in computers with applications from Shanxi University, Taiyuan, China, in 2005 and 2011, respectively. He is currently a Professor with the Key Laboratory of Computational Intelligence and Chinese Information Processing, Ministry of Education, Shanxi University. He is best known for multigranulation rough sets in learning from categorical data and granular computing. He is involved in research on pattern recognition, feature selection, rough set theory, granular computing, and artificial intelligence. He has published over 80 articles on these topics in international journals. He served on the Editorial Board of the International Journal of Knowledge-Based Organizations and Artificial Intelligence Research. He has served as the Program Chair or Special Issue Chair of the Conference on Rough Sets and Knowledge Technology, the Joint Rough Set Symposium, and the Conference on Industrial Instrumentation and Control, and a PC member of many machine learning, data mining, and granular computing conferences.