# Intuitionistic Fuzzy Rough Set-Based Granular Structures and Attribute Subset Selection

Anhui Tan, Wei-Zhi Wu, Yuhua Qian, Jiye Liang, Jinkun Chen, and Jinjin Li

***Abstract*—Attribute subset selection is an important issue in data mining and information processing. However, most automatic methodologies consider only the relevance factor between samples while ignoring the diversity factor. This may not allow the utilization value of hidden information to be exploited. For this reason, we propose a hybrid model named intuitionistic fuzzy (IF) rough set to overcome this limitation. The model combines the technical advantages of rough set and IF set and can effectively consider the above-mentioned statistical factors. First, fuzzy information granules based on IF relations are defined and used to characterize the hierarchical structures of the lower and upper approximations of IF rough set within the framework of granular computing. Then, the computation of IF rough approximations and knowledge reduction in IF information systems are investigated. Third, based on the approximations of IF rough set, significance measures are developed to evaluate the approximation quality and classification ability of IF relations. Furthermore, a forward heuristic algorithm for finding one optimal reduct of IF information systems is developed using these measures. Finally, numerical experiments are conducted on public datasets to examine the effectiveness and efficiency of the proposed algorithm in terms of the number of selected attributes, computational time, and classification accuracy.**

***Index Terms*—Attribute reduction, granular structure, intuitionistic fuzzy (IF) relation, IF rough set, rough approximation.**

Manuscript received April 2, 2018; revised June 27, 2018; accepted July 24, 2018. Date of publication August 2, 2018; date of current version February 27, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61602415, Grant 61573321, Grant 41631179, Grant 61672332, Grant 61773247, and Grant 61773349, in part by the Program for the Outstanding Innovative Teams of Higher Learning Institutions of Shanxi, the Program for the Young San Jin Scholars of Shanxi, and in part by the Natural Science Foundation of Zhejiang Province of China under Grant LY18F030017.
*(Corresponding author: Anhui Tan.)*

A. Tan is with the School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhoushan 316022, China, and also with the School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China (e-mail: tananhui86@163.com).

W.-Z. Wu is with the School of Mathematics, Physics and Information Science, Zhejiang Ocean University, Zhoushan 316022, China, and also with the Key Laboratory of Oceanographic Big Data Mining and Application of Zhejiang Province, Zhoushan 316022, China (e-mail: wuwz@zjou.edu.cn).

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

J. Liang is with the School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China, and also with the Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, China (e-mail: ljy@sxu.edu.cn).

J. Chen and J. Li are with the School of Mathematics and Statistics, Minnan Normal University, Zhangzhou 363000, China (e-mail: cjk99@163.com; jinjinli@mnnu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2018.2862870

## I. Introduction

Rough set theory and fuzzy set theory are two important tools for conceptualizing and analyzing various types of data. They have attracted considerable attention in the domains of granular computing and machine learning. Pawlak's rough set [1] utilizes an indiscernibility relation to construct the lower and upper approximations of any arbitrary subset of the universe. It can directly handle categorical attributes [2]–[5], whereas numerical and continuous ones must be discretized into several intervals, which are assigned a set of symbolic values before the rough set model is applied. Different discretization methods may change the original distribution of data and lead to information loss. In view of this observation, fuzzy rough set [6], [7], as a combination of fuzzy set and rough set, has proven to be an effective tool for overcoming the puzzle of discretization and can be used to analyze numerical and continuous attributes without additional preprocessing steps. Over the past decade, this theory has been successfully applied in gene expression, cluster analysis, feature selection, and rule extraction [8]–[11].

We elaborate the advantages of fuzzy rough set from the following two angles. First, this model [12]–[16] was introduced to express the idea of granular computing, which refers to the theories and methodologies that utilize granules in information processing. Zadeh [16] first proposed and discussed the issue of fuzzy information granulation, which is used to generate a family of fuzzy information granules from numerical and continuous data. Pedrycz *et al.* [17] presented a granular evaluation of fuzzy models and delivered an augmentation of fuzzy models by forming information granules. Qian *et al.* [18] developed a measure, named a fuzzy granular structure distance, to discriminate the difference between fuzzy granular structures. By considering the intention and extension of information, Xu and Li [19] addressed the mechanism required to train arbitrary fuzzy information granules. Chen *et al.* [20] introduced the concept of granular fuzzy set based on fuzzy similarity relations and described the granular structures of fuzzy rough approximation operators. Second, fuzzy rough set is widely utilized for feature selection (also called attribute reduction) [21], [22]. Jensen and Shen [10], [23] first defined a notion of dependence function-based reduct and designed a heuristic reduction algorithm using a fuzzy rough set. In [24], a computational domain was presented to improve the computational efficiency of the algorithm proposed in [10]. In [25], the authors examined the granular structures of lower and upper fuzzy rough approximations and constructed a discernibility matrix, using which the foundation of fuzzy-rough attribute reduction was established.

Considerable effort has been devoted to developing reduction methods by remodeling the discernibility matrices [26], [27]. Wang *et al.* [28]–[30] introduced the concept of fuzzy decision of samples for feature selection using the fuzzy rough set. It is noteworthy that the fuzzy rough set is sensitive to noise interference. For this reason, the robustness is enhanced to improve the generalizability of fuzzy rough set models in the presence of attribute noise and class noise [31]–[33]. The perspective of an additional class of attribute reduction algorithms with the fuzzy rough set is fuzzy information entropy. Hu *et al.* [34] employed information entropy to redefine the definitions of attribute subset reduct and relative reduct. Zhang *et al.* [35] presented the granular structures of fuzzy rough set under general fuzzy relations and proposed granule-based information entropy for feature selection.

The intuitionistic fuzzy (IF) set [36], [37], as an intuitive generalization of fuzzy set, simultaneously takes the membership and nonmembership of objects belonging to target sets into account. This model possesses a stronger ability to express the delicate ambiguities of information in the objective world than do the traditional fuzzy set. The combination of the IF set and rough set resulted in the introduction of IF rough set [38]–[40]. It employs the rough set approximations to describe the IF set by associating a collection of IF relations between objects. Zhou *et al.* [41], [42] examined approximation operators of the IF rough set by using constructive and axiomatic approaches. Huang *et al.* [43], [44] developed IF rough set models for interval-valued and multigranulation data and discussed the hierarchical structures and uncertainty measures of IF information [45]. Zhang *et al.* [46] provided a general framework of IF rough set models related to general binary relations between two universes. Hesameddini *et al.* [47] and Zhang [48] investigated the attribute reduction of the IF rough set and constructed the reduction structures based on discernibility matrices. The existence of all these results indicates that the IF rough set has received wide attention in the research community in recent years. However, the studies were focused mainly on model generalization [49]–[51], property exploration [52]–[54], and measure description [55], [56], while the advantages of the IF rough set have rarely been demonstrated in different fields, such as granular computing and feature selection.

The fuzzy rough set considers only the relevance of samples while overlooking the diversity. To overcome this limitation, we in this paper employ the IF rough set model to construct the fuzzy lower and upper approximations of decisions. This model aims to maintain both the maximal degrees of samples' membership to their own classes and nonmembership to other classes simultaneously. We first define several types of fuzzy information granules and then present the hierarchical structures of the IF rough set from the viewpoint of granular computing. The significance measures are introduced to evaluate the classification ability of the IF relations in an IF decision system. A forward reduction algorithm is further proposed for finding a reduct of the system. Finally, we compare the proposed algorithm with existing algorithms. The experimental results show that the new algorithm is effective and efficient in the aspects of dimensionality reduction, computational time, and classification accuracy.

The rest of this paper is organized as follows. In Section I, we briefly review some basic concepts of IF set, IF relation, and IF rough set. In Section II, we define several types of fuzzy granules and characterize the hierarchical structures of the lower and upper approximations of the IF rough set by using these granules. In Section III, we present a simple way to compute the IF rough approximations in the IF decision systems. In Section IV, we discuss the knowledge reduction of IF decision information systems in detail. A dependence function-based heuristic reduction algorithm is then formulated to compute a suboptimal reduct of IF decision systems. In Section V, numerical experiments to verify the feasibility and stability of the proposed algorithm are described. The paper is concluded with a summary.

### A. Preliminaries

In this section, we review several basic concepts related to the IF set and IF relation. The notion of the IF rough set is then presented. More details can be found in [36], [37], [40], and [41].

#### 1) IF Set and IF Relation:

*Definition 1 ([36], [37]):* Let $U$ be the universe. An IF set $X$ on $U$ is

$$X = \left\{ \frac{(\mu_X(x), \gamma_X(x))}{x} | x \in U \right\}$$

where $\mu_X : U \to [0, 1]$ and $\gamma_X : U \to [0, 1]$ satisfy $\mu_X(x) + \gamma_X(x) \leq 1$ for any $x \in U$. $\mu_X(x)$ and $\gamma_X(x)$ are called the membership and nonmembership degrees of object $x$ to set $X$, respectively. The pair $(\mu_X(x), \gamma_X(x))$ is an IF number of $x$ w.r.t. $X$.

The family of all IF sets on $U$ is denoted by $\mathcal{IF}$, in which $\widetilde{1} = \{\frac{(1,0)}{x} | x \in U\}$ is the IF universal set and $\widetilde{\emptyset} = \{\frac{(0,1)}{x} | x \in U\}$ is the IF empty set. Obviously, each ordinary fuzzy set $X$ can be represented by an IF set: $X = \{\frac{(\mu_X(x), 1-\mu_X(x))}{x} | x \in U\}$.

Denote the quality of $X$ as $|X| = \sum_{x \in U} \frac{1+\mu_X(x)-\gamma_X(x)}{2}$ [57], [58], where 1 is a translation factor that ensures the value is a positive number, and quotient 2 is a scaling factor that ensures the value is bounded within 0 and 1. Moreover, define the probability quality of $X$ w.r.t. the universe as $p(X) = \frac{|X|}{|U|}$. We obtain the following basic properties.

*Property 1:* Let $U$ be the universe and $X, Y \in \mathcal{IF}$.

1) $0 \leq |X| \leq |U|$ and $0 \leq p(X) \leq 1$.
2) $p(X) = 1$ iff $X = \widetilde{1}$.
3) $p(X) = 0$ iff $X = \widetilde{\emptyset}$.
4) If $Y \subseteq X$, then $|Y| \leq |X|$ and $p(Y) \leq p(X)$.

We can see from Property 1 that the smaller the membership and the larger the nonmembership degree, the smaller the IF set. In extreme cases, $\widetilde{1}$ and $\widetilde{\emptyset}$ are the largest and smallest IF sets, respectively. All IF sets in $\mathcal{IF}$ on the universe generate a complete lattice.

*Definition 2 ([59]):* Let $U$ be the universe. An IF binary relation $R$ on $U$ is defined as

$$R = \{< (x, y), \mu_R(x, y), \gamma_R(x, y) > | (x, y) \in U \times U\}$$

where $\mu_R : U \times U \to [0, 1]$ and $\gamma_R : U \times U \to [0, 1]$ satisfy $\mu_R(x, y) + \gamma_R(x, y) \leq 1$ for any $(x, y) \in U \times U$. $\mu_R(x, y)$ and

$\gamma_R(x, y)$ are, respectively, the similarity and diversity degrees of $x$ to $y$.

For two IF relations $R$ and $B$, we say $R$ is finer than $B$, denoted by $R \preceq B$, if and only if $\mu_R(x, y) \leq \mu_B(x, y)$ and $\gamma_R(x, y) \geq \gamma_B(x, y)$ for any $(x, y) \in U \times U$.

Assume $U = \{x_1, x_2, \ldots, x_n\}$. An IF relation is then represented by a matrix $R = [(\mu_R(x_i, x_j), \gamma_R(x_i, x_j))]_{n\times n}$, where the $(i, j)$th entry displays the IF number between $x_i$ and $x_j$.

An IF relation $R$ is an IF tolerance relation if it satisfies the following:

a) reflexive: $\mu_R(x, x) = 1, \gamma_R(x, x) = 0$;

b) symmetric: $\mu_R(x, y) = \mu_R(y, x), \gamma_R(x, y) = \gamma_R(x, y)$.

The representative matrix of an IF tolerance relation is symmetric whose diagonal values are $\widetilde{1}$. Throughout this study, we limit the IF relations within IF tolerance ones.

*2) IF Rough Set:*

*Definition 3 ([41], [42]):* Let $U$ be the universe and $R$ an IF relation. For any IF set $X \in \mathcal{IF}$, the lower and upper approximations of $X$ w.r.t. $R$ are, respectively, defined as

$$\underline{R}(X) = \left\{ \frac{(\mu_{\underline{R}(X)}(x), \gamma_{\underline{R}(X)}(x))}{x} | x \in U \right\}$$

$$\overline{R}(X) = \left\{ \frac{(\mu_{\overline{R}(X)}(x), \gamma_{\overline{R}(X)}(x))}{x} > | x \in U \right\}$$

where

$$\mu_{\underline{R}(X)}(x) = \inf_{y\in U} \max(\gamma_R(x, y), \mu_X(y))$$

$$\gamma_{\underline{R}(X)}(x) = \sup_{y\in U} \min(\mu_R(x, y), \gamma_X(y))$$

$$\mu_{\overline{R}(X)}(x) = \sup_{y\in U} \min(\mu_R(x, y), \mu_X(y))$$

$$\gamma_{\overline{R}(X)}(x) = \inf_{y\in U} \max(\gamma_R(x, y), \gamma_X(y)).$$

The pairs $(\mu_{\underline{R}(X)}(x), \gamma_{\underline{R}}(X)(x))$ and $(\mu_{\overline{R}(X)}(x), \gamma_{\overline{R}(X)}(x))$ are the IF numbers of object $x$ to the lower and upper approximation sets, respectively. If $\underline{R}(X) = \overline{R}(X)$, then $(\underline{R}(X), \overline{R}(X))$ is referred to as an IF definable set; otherwise, as an IF rough set.

*Property 2:* Let $R$ and $B$ be two IF relations. If $R \preceq B$, then for any $X \in \mathcal{IF}$, it holds that

a) $\underline{R}(X) \subseteq X \subseteq \overline{R}(X)$;

b) $\underline{B}(X) \subseteq \underline{R}(X)$;

c) $\overline{R}(X) \subseteq \overline{B}(X)$.

Property 2 discloses that IF approximations are monotonic with the increasing or decreasing in the granularity of the IF relations. A finer IF relation leads to a larger lower approximation and a smaller upper approximation w.r.t. a given IF set.

*3) IF Decision Information System:* Classification is one of the most important tasks in the machine learning field. For dealing with IF information and IF information-related classification, in this section, we formally define the notion of IF information systems and examine IF rough approximations in this context.

*Definition 4:* Let $U$ be the universe and $\mathcal{R}$ a family of IF relations. The pair $S = (U, \mathcal{R})$ is called an IF information system. Moreover, if $d$ is a decision attribute, then $S = (U, \mathcal{R} \cup \{d\})$ is called an IF decision system.

For processing real datasets, it is possible to construct IF relations from the attributes by applying various methods. Then, the datasets can be transformed to IF information systems for further analysis.

Assume $d$ is a decision attribute, which divides the universe into a family of decision classes $U/d = \{d_1, d_2, \ldots, d_l\}$ according to different decision labels. Each decision class can be typically written as an IF set: $d_i = \{\frac{(\mu_{d_i}(x), \gamma_{d_i}(x))}{x} | x \in U\}$, where

$$(\mu_{d_i}(x), \gamma_{d_i}(x)) = \begin{cases} (1, 0), & x \in d_i; \\ (0, 1), & x \notin d_i. \end{cases}$$

Data often contain redundant information, which is irrelevant for knowledge representation and decision classification. Its removal can save the storage space and computational time, and raise learning efficiency. In this section, we describe the use of the IF rough set to reduce irrelevant IF relations and select important ones from IF decision systems.

Given a subset of IF relations $\mathcal{B} \subseteq \mathcal{R}$. A special IF relation may be induced from $\mathcal{B}$ by taking the interrelationship among the IF relations. We represent the corresponding IF relation w.r.t. $\mathcal{B}$ by the symbol $\mathbf{B}$. Note that the IF relation induced varies with different constructive methods. Based on this generalized definition, we examine the knowledge reduction of IF decision systems in the following.

Let $S = (U, \mathcal{R} \cup \{d\})$ be an IF decision system with $U/d = \{d_1, d_2, \ldots, d_l\}$. Given a subset $\mathcal{B} \subseteq \mathcal{R}$, we can compute the distributions of IF lower and upper approximations of decision classes as

$$\underline{\mathbf{B}}(d) = \{\underline{\mathbf{B}}(d_1), \underline{\mathbf{B}}(d_2), \ldots, \underline{\mathbf{B}}(d_l))\}$$

$$\overline{\mathbf{B}}(d) = \{\overline{\mathbf{B}}(d_1), \overline{\mathbf{B}}(d_2), \ldots, \overline{\mathbf{B}}(d_l))\}.$$

*Definition 5:* Let $S = (U, \mathcal{R} \cup \{d\})$ be an IF decision system with $U/d = \{d_1, d_2, \ldots, d_l\}$ and $\mathcal{B} \subseteq \mathcal{R}$. Define $\text{Pos}_{\mathcal{B}}(d) = \cup_{i=1}^{l} \underline{\mathbf{B}}(d_i)$ as the IF positive region w.r.t. $\mathcal{B}$. $\mathcal{B}$ is called a reduct of $S$ if it satisfies the following:

a) $\text{Pos}_{\mathcal{B}}(d) = \text{Pos}_{\mathcal{R}}(d)$

b) $\text{Pos}_{\mathcal{B}-\{R\}}(d) \subset \text{Pos}_{\mathcal{B}}(d) \;\; \forall R \in \mathcal{B}$.

In Definition 5, the IF positive region is an IF set generated by taking the union of the lower approximations of all decision classes. It reflects the classification ability of any subset $\mathcal{B}$. The larger the IF positive region, the stronger the classification ability of $\mathcal{B}$.

## II. Granular Structures of Intuitionistic Fuzzy Approximations

In this section, we examine the granular structures of IF rough approximation operators. We first define several types of fuzzy granules and clarify the hierarchies of the lower and upper approximations of IF set using these granules.

*Definition 6:* Let $U$ be the universe and $R$ an IF relation. Given a real number $\lambda \in [0, 1]$ and $x \in U$, we define a type-1 fuzzy granule as

$$[x^{\lambda}_{\gamma_R}](y) = \begin{cases} \lambda, & \gamma_R(x,y) < \lambda \\ 0, & \gamma_R(x,y) \geq \lambda \end{cases}.$$

In Definition 6, the granule $[x^{\lambda}_{\gamma_R}]$ is based on the diversity function $\gamma_R$ of the relation $R$. Parameter $\lambda$ influences the size of the fuzzy granule: the larger is the value of $\lambda$, the larger is the fuzzy granule. The lower approximations can be hierarchically described as follows.

*Theorem 1:* Let $U$ be the universe and $R$ an IF relation. For $X \in \mathcal{IF}$ and $x \in U$, we have

$$\mu_{\underline{R}(X)}(x) = \sup\{\lambda | [x^{\lambda}_{\gamma_R}] \subseteq \mu_X\}.$$

*Proof:*

1) For any $x \in U$, let $\mu_{\underline{R}(X)}(x) = \lambda_0$. Then, from Definition 3, it holds that $\max(\gamma_R(x,y), \mu_X(y)) \geq \lambda_0$ for any $y \in U$, and there exists one $y_0 \in U$ satisfying $\max(\gamma_R(x,y_0), \mu_X(y_0)) = \lambda_0$.
We have that if $\gamma_R(x,y) < \lambda_0$, then $\mu_X(y) \geq \lambda_0$ for any $y \in U$. This means that $[x^{\lambda_0}_{\gamma_R}](y) \leq \mu_X(y)$ for any $\gamma_R(x,y) < \lambda_0$. On the other hand, if $\gamma_R(x,y) \geq \lambda_0$, then $[x^{\lambda_0}_{\gamma_R}](y) = 0$. On the whole, $[x^{\lambda_0}_{\gamma_R}] \subseteq \mu_X$.
2) Assume that there is some $\lambda_1 > \lambda_0$ such that $[x^{\lambda_1}_{\gamma_R}] \subseteq \mu_X$. It follows that $[x^{\lambda_1}_{\gamma_R}](y) \leq \mu_X(y)$ for any $y \in U$. We know from Definition 6 that the condition $\gamma_R(x,y) < \lambda_1$ implies $[x^{\lambda_1}_{\gamma_R}](y) = \lambda_1$. With the fact of $[x^{\lambda_1}_{\gamma_R}] \subseteq \mu_X$, we have that if $\gamma_R(x,y) < \lambda_1$, then $\mu_X(y) \geq \lambda_1$. Subsequently

$$\begin{aligned} \mu_{\underline{R}(X)}(x) &= \inf_{y\in U} \max(\gamma_R(x,y), \mu_X(y)) \\ &= \inf_{\gamma_R(x,y)<\lambda_1} \max(\gamma_R(x,y), \mu_X(y)) \\ &\geq \min\{\lambda_1, \lambda_1\} = \lambda_1. \end{aligned}$$

This implies that $\mu_{\underline{R}(X)}(x) \geq \lambda_1 > \lambda_0$, which is in contradiction with $\mu_{\underline{R}(X)}(x) = \lambda_0$. Hence, $\mu_{\underline{R}(X)}(x) = \sup\{\lambda | [x^{\lambda}_{\gamma_R}] \subseteq \mu_X, \lambda \in [0,1]\}$. We complete the proof. ∎

From Theorem 1, one can see that the membership of an object $x$ belonging to the set $\mu_{\underline{R}(X)}$ is the maximal real number, which guarantees that the granule $[x^{\lambda}_{\gamma_R}]$ is contained in $\mu_X$. An example is used to examine this idea.

*Example 1:* Continued from Example 1.

In Example 1, we have obtained that $\mu_{\underline{R}(X)} = \frac{0.5}{x_1} + \frac{0.5}{x_2} + \frac{0.2}{x_3}$. From Definition 6, we have that

$$[x^{0.5}_{1\gamma_R}] = \frac{0.5}{x_1} + \frac{0.5}{x_2} + \frac{0}{x_3}$$

$$[x^{0.5}_{2\gamma_R}] = \frac{0.5}{x_1} + \frac{0.5}{x_2} + \frac{0}{x_3}$$

$$[x^{0.2}_{3\gamma_R}] = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.2}{x_3}.$$

Since $\mu_X = \frac{0.5}{x_1} + \frac{0.6}{x_2} + \frac{0.2}{x_3}$, we can see that $[x^{0.5}_{1\gamma_R}] \subseteq \mu_X$, $[x^{0.6}_{2\gamma_R}] \subseteq \mu_X$ and $[x^{0.2}_{3\gamma_R}] \subseteq \mu_X$.

Moreover, if we take $\lambda > 0.5$, it holds $[x^{\lambda}_{1\gamma_R}](x_1) = \lambda > \mu_X(x_1)$ and $[x^{\lambda}_{2\gamma_R}](x_2) = \lambda > \mu_X(x_2)$, which follows the conclusion that $[x^{\lambda}_{1\gamma_R}] \subseteq \mu_X$ and $[x^{\lambda}_{2\gamma_R}] \subseteq \mu_X$ do not hold. If $\lambda > 0.2$, it follows that $[x^{\lambda}_{3\gamma_R}](x_3) = \lambda > \mu_X(x_3)$, which follows that $[x^{\lambda}_{3\gamma_R}] \subseteq \mu_X$ does not hold.

By these analyses, we assert that $\mu_{\underline{R}(X)} = \frac{0.5}{x_1} + \frac{0.5}{x_2} + \frac{0.2}{x_3}$. This is in accordance with Theorem 1.

*Definition 7:* Let $U$ be the universe and $R$ an IF relation. Given a real number $\lambda \in [0,1]$ and $x \in U$, define a type-2 fuzzy granule as

$$[x^{\lambda}_{\mu_R}](y) = \begin{cases} \lambda, & \mu_R(x,y) > \lambda; \\ 1, & \mu_R(x,y) \leq \lambda. \end{cases}$$

The granule $[x^{\lambda}_{\mu_R}]$ is based on the similarity function $\mu_R$ of the relation $R$ and parameter $\lambda$: the larger is the value of $\lambda$, the larger is the fuzzy granule. We have the following theorem.

*Theorem 2:* Let $U$ be the universe and $R$ an IF relation. For $X \in \mathcal{IF}$ and $x \in U$, we have

$$\gamma_{\underline{R}(X)}(x) = \inf\{\lambda | [x^{\lambda}_{\mu_R}] \supseteq \gamma_X\}.$$

*Proof:*

1) For any $x \in U$, let $\gamma_{\underline{R}(X)}(x) = \lambda_0$. Then, from Definition 3, it holds that $\min(\mu_R(x,y), \gamma_X(y)) \leq \lambda_0$ for any $y \in U$, and there exists one $y_0 \in U$ satisfying that $\min(\mu_R(x,y_0), \gamma_X(y_0)) = \lambda_0$.
We have that if $\mu_R(x,y) > \lambda_0$, then $\gamma_X(y) \leq \lambda_0$ for any $y \in U$. This means that $[x^{\lambda_0}_{\mu_R}](y) \geq \gamma_X(y)$ for $\mu_R(x,y) > \lambda_0$. On the other hand, if $\mu_R(x,y) \leq \lambda_0$, then $[x^{\lambda_0}_{\mu_R}](y) = 1$. On the whole, $[x^{\lambda_0}_{\mu_R}] \supseteq \gamma_X$.
2) Assume that there is some $\lambda_1 < \lambda_0$ such that $[x^{\lambda_1}_{\mu_R}] \supseteq \gamma_X$. It follows that $[x^{\lambda_1}_{\mu_R}](y) \geq \gamma_X(y)$ for any $y \in U$. We know from that the condition $\mu_R(x,y) > \lambda_1$ implies $[x^{\lambda_1}_{\mu_R}](y) = \lambda_1$. With the fact of $[x^{\lambda_1}_{\mu_R}] \supseteq \gamma_X$ and Definition 7, we have that if $\mu_R(x,y) > \lambda_1$, then $\gamma_X(y) \leq \lambda_1$. Subsequently

$$\begin{aligned} \gamma_{\underline{R}(X)}(x) &= \sup_{y\in U} \min(\mu_R(x,y), \gamma_X(y)) \\ &= \sup_{\mu_R(x,y)>\lambda_1} \min(\mu_R(x,y), \gamma_X(y)) \\ &\leq \max\{\lambda_1, \lambda_1\} = \lambda_1. \end{aligned}$$

This implies that $\gamma_{\underline{R}(X)}(x) \leq \lambda_1 < \lambda_0$, which is in contradiction with $\gamma_{\underline{R}(X)}(x) = \lambda_0$. Hence, $\gamma_{\underline{R}(X)}(x) = \inf\{\lambda | [x^{\lambda}_{\mu_R}] \supseteq \gamma_X\}$. We complete the proof. ∎

From Theorem 2, one can see that the membership of an object $x$ to the set $\gamma_{\underline{R}(X)}$ is the minimal real number, which guarantees that the granule $[x^{\lambda}_{\mu_R}]$ contains $\gamma_X$. An example is used to illustrate this idea.

*Example 2:* Continued from Example 1.

Example 1 indicates that $\gamma_{\underline{R}(X)} = \frac{0.3}{x_1} + \frac{0.4}{x_2} + \frac{0.7}{x_3}$. From Definition 7, we have that

$$[x^{0.3}_{1\mu_R}] = \frac{0.3}{x_1} + \frac{0.3}{x_2} + \frac{1}{x_3}$$

$$[x^{0.4}_{2\mu_R}] = \frac{0.4}{x_1} + \frac{0.4}{x_2} + \frac{1}{x_3}$$

$$[x^{0.7}_{3\mu_R}] = \frac{1}{x_1} + \frac{1}{x_2} + \frac{0.7}{x_3}.$$

Since $\gamma_X = \frac{0.3}{x_1} + \frac{0.1}{x_2} + \frac{0.7}{x_3}$, we can see that $[x^{0.3}_{1\gamma_R}] \supseteq \gamma_X$, $[x^{0.4}_{2\gamma_R}] \supseteq \gamma_X$, and $[x^{0.7}_{3\gamma_R}] \supseteq \gamma_X$.

Moreover, if $\lambda < 0.3$, it holds $[x^{\lambda}_{1\mu_R}](x_3) = \lambda < \gamma_X(x_3)$, which follows that $[x^{\lambda}_{1\mu_R}] \supseteq \gamma_X$ does not hold. If $\lambda < 0.4$, we have that $[x^{\lambda}_{2\mu_R}](x_3) = \lambda < \gamma_X(x_3)$, then $[x^{\lambda}_{2\mu_R}] \supseteq \gamma_X$ does not hold. If $\lambda < 0.7$, it follows $[x^{\lambda}_{3\mu_R}](x_3) = \lambda < \gamma_X(x_3)$. Thus, $[x^{\lambda}_{3\mu_R}] \supseteq \gamma_X$ does not hold. This is in accordance with the conclusion in Theorem 2.

From the examples mentioned above, we see that the IF lower approximations can be characterized and computed using fuzzy information granules. We next examine the IF upper approximations. The following theorems can be similarly verified.

*Theorem 3:* $\mu_{\overline{R}(X)}(x) = \inf\{\lambda | [x^{\lambda}_{\mu_R}] \supseteq \mu_X\}$.

*Proof:* The proof is similar to that of Theorem 2. ■

*Example 3:* Continued from Example 1.

In Example 1, we have that $\mu_{\overline{R}(X)} = \frac{0.6}{x_1} + \frac{0.6}{x_2} + \frac{0.4}{x_3}$. We now use Theorem 3 to examine $\mu_{\overline{R}(X)}$.

From Definition 7, we have that

$$[x^{0.6}_{1\mu_R}] = \frac{0.6}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}$$

$$[x^{0.6}_{2\mu_R}] = \frac{1}{x_1} + \frac{0.6}{x_2} + \frac{1}{x_3}$$

$$[x^{0.4}_{3\mu_R}] = \frac{1}{x_1} + \frac{1}{x_2} + \frac{0.4}{x_3}.$$

Since $\mu_X = \frac{0.5}{x_1} + \frac{0.6}{x_2} + \frac{0.2}{x_3}$, we can see that $[x^{0.6}_{1\mu_R}] \supseteq \mu_X$, $[x^{0.6}_{2\mu_R}] \supseteq \mu_X$, and $[x^{0.6}_{3\mu_R}] \supseteq \mu_X$.

Moreover, if we take $\lambda < 0.6$, it holds $[x^{\lambda}_{1\mu_R}](x_2) = \lambda < \mu_X(x_2)$ and $[x^{\lambda}_{2\mu_R}](x_1) = \lambda < \mu_X(x_2)$, which follows that $[x^{\lambda}_{1\mu_R}] \supseteq \mu_X$ and $[x^{\lambda}_{2\gamma_R}] \supseteq \mu_X$ do not hold. If $\lambda < 0.4$, it holds $[x^{\lambda}_{3\gamma_R}](x_2) = \lambda < \mu_X(x_2)$, which follows that $[x^{\lambda}_{3\mu_R}] \supseteq \mu_X$ does not hold.

Theorem 3 implies that $\mu_{\overline{R}(X)}(x_1) = 0.6$, $\mu_{\overline{R}(X)}(x_2) = 0.6$, and $\mu_{\overline{R}(X)}(x_3) = 0.4$. Consequently, $\mu_{\overline{R}(X)} = \frac{0.6}{x_1} + \frac{0.6}{x_2} + \frac{0.4}{x_3}$.

*Theorem 4:* $\gamma_{\overline{R}(X)}(x) = \sup\{\lambda | [x^{\lambda}_{\gamma_R}] \subseteq \gamma_X\}$.

*Proof:* The proof is similar to that of Theorem 1. ■

*Example 4:* Continued from Example 1.

From Example 1, we obtain that $\gamma_{\overline{R}(X)} = \frac{0.2}{x_1} + \frac{0.1}{x_2} + \frac{0.5}{x_3}$.

From Definition 6, we compute that

$$[x^{0.2}_{1\gamma_R}] = \frac{0.2}{x_1} + \frac{0}{x_2} + \frac{0}{x_3}$$

$$[x^{0.1}_{2\gamma_R}] = \frac{0}{x_1} + \frac{0.1}{x_2} + \frac{0}{x_3}$$

$$[x^{0.5}_{3\gamma_R}] = \frac{0}{x_1} + \frac{0}{x_2} + \frac{0.5}{x_3}.$$

Since $\gamma_X = \frac{0.3}{x_1} + \frac{0.1}{x_2} + \frac{0.7}{x_3}$, we can see that $[x^{0.2}_{1\gamma_R}] \subseteq \gamma_X$, $[x^{0.1}_{2\gamma_R}] \subseteq \mu_X$, and $[x^{0.5}_{3\gamma_R}] \subseteq \mu_X$.

If $\lambda > 0.2$, it holds $[x^{\lambda}_{1\gamma_R}](x_2) = \lambda > \gamma_X(x_2)$, which follows that $[x^{\lambda}_{1\gamma_R}] \subseteq \gamma_X$ does not hold. If $\lambda > 0.1$, we have $[x^{\lambda}_{2\gamma_R}](x_2) = \lambda > \gamma_X(x_2)$. This indicates that $[x^{\lambda}_{2\gamma_R}] \subseteq \gamma_X$ does not hold. If $\lambda > 0.5$, it holds $[x^{\lambda}_{3\gamma_R}](x_2) = \lambda > \gamma_X(x_2)$, which follows that $[x^{\lambda}_{3\gamma_R}] \subseteq \gamma_X$ does not hold.

Theorem 4 indicates that $\gamma_{\overline{R}(X)}(x_1) = 0.6$, $\mu_{\overline{R}(X)}(x_2) = 0.6$, and $\mu_{\overline{R}(X)}(x_3) = 0.4$. Consequently, $\mu_{\overline{R}(X)} = \frac{0.6}{x_1} + \frac{0.6}{x_2} + \frac{0.4}{x_3}$.

In this section, we presented the construction of two types of fuzzy granules based on IF relations, and then, the hierarchical structures of IF rough lower and upper approximations were both characterized. In this sense, the lower and upper approximations of the IF set can be computed by the granular computing-based methods.

## III. Computation of Intuitionistic Fuzzy Approximations

The computation of IF approximations is a basic issue for the further application of the IF rough set. We employ the following example to illustrate the acquisition of the rough approximations of the IF set.

*Example 5:* Assume $U = \{x_1, x_2, x_3\}$ and the IF relation $R$ is

$$R = \begin{pmatrix} (1,0) & (0.6,0.2) & (0.3,0.6) \\ (0.6,0.2) & (1,0) & (0.4,0.5) \\ (0.3,0.6) & (0.4,0.5) & (1,0) \end{pmatrix}.$$

Let $X = \frac{(0.5,0.3)}{x_1} + \frac{(0.6,0.1)}{x_2} + \frac{(0.2,0.7)}{x_3}$. We now compute the lower and upper approximations of $X$.

We have that

$$\mu_{\underline{R}(X)}(x_1) = \inf\{\max(0,0.5), \max(0.2,0.6), \max(0.6,0.2)\} = 0.5$$

$$\mu_{\underline{R}(X)}(x_2) = \inf\{\max(0.2,0.5), \max(0,0.6), \max(0.5,0.2)\} = 0.5$$

$$\mu_{\underline{R}(X)}(x_3) = \inf\{\max(0.6,0.5), \max(0.5,0.6), \max(0,0.2)\} = 0.2,$$

and

$$\gamma_{\underline{R}(X)}(x_1) = \sup\{\min(1,0.3), \min(0.6,0.1), \min(0.3,0.7)\} = 0.3$$

$$\gamma_{\underline{R}(X)}(x_2) = \sup\{\min(0.6,0.3), \min(1,0.1), \min(0.4,0.7)\} = 0.4$$

$$\gamma_{\underline{R}(X)}(x_3) = \sup\{\min(0.3,0.3), \min(0.4,0.1), \min(1,0.7)\} = 0.7.$$

This means that $\underline{R}(X) = \frac{(0.5,0.3)}{x_1} + \frac{(0.5,0.4)}{x_2} + \frac{(0.2,0.7)}{x_3}$.

Moreover,

$$\mu_{\overline{R}(X)}(x_1) = \sup\{\min(1,0.5), \min(0.6,0.6), \min(0.3,0.2)\} = 0.6$$

$$\mu_{\overline{R}(X)}(x_2) = \sup\{\min(0.6,0.5), \min(1,0.6), \min(0.4,0.2)\} = 0.6$$

$$\mu_{\overline{R}(X)}(x_3) = \sup\{\min(0.3,0.5), \min(0.4,0.6), \min(1,0.2)\} = 0.4.$$

Also

$$\gamma_{\overline{R}(X)}(x_1) = \inf\{\max(0, 0.3), \max(0.2, 0.1), \max(0.6, 0.7)\} = 0.2$$

$$\gamma_{\overline{R}(X)}(x_2) = \inf\{\max(0.2, 0.3), \max(0, 0.1), \max(0.5, 0.7)\} = 0.1$$

$$\gamma_{\overline{R}(X)}(x_3) = \inf\{\max(0.6, 0.3), \max(0.5, 0.1), \max(0, 0.7)\} = 0.5.$$

Subsequently, $\overline{R}(X) = \frac{(0.6,0.2)}{x_1} + \frac{(0.6,0.1)}{x_2} + \frac{(0.4,0.5)}{x_3}$.

We clearly see that $\underline{R}(X) \subseteq X \subseteq \overline{R}(X)$.

The computation of IF approximations is a difficult task because it requires a full run through all the IF numbers in the IF relation. To facilitate this process, we examine the special properties of IF approximations in IF decision systems. We arrive at the following statements for the IF rough approximations of decision classes.

*Theorem 5:* Let $S = (U, \mathcal{R} \cup \{d\})$ be an IF decision system with $R \in \mathcal{R}$ and $U/d = \{d_1, d_2, \ldots, d_l\}$. For any $1 \le i \le l$, we have

$$\mu_{\underline{R}(d_i)}(x) = \inf_{y \notin d_i} \gamma_R(x, y)$$

$$\gamma_{\underline{R}(d_i)}(x) = \sup_{y \notin d_i} \mu_R(x, y)$$

$$\mu_{\overline{R}(d_i)}(x) = \sup_{y \in d_i} \mu_R(x, y)$$

$$\gamma_{\overline{R}(d_i)}(x) = \inf_{y \in d_i} \gamma_R(x, y).$$

*Proof:*

1) From Definition 3, we have that $\mu_{\underline{R}(d_i)}(x) = \inf_{y \in U} \max(\gamma_R(x, y), \mu_{d_i}(y))$. If $y \in d_i$, then $\mu_{d_i}(y) = 1$, which leads to that $\max(\gamma_R(x, y), \mu_{d_i}(y)) = \max(\gamma_R(x, y), 1) = 1$. Consequently, $\mu_{\underline{R}(d_i)}(x) = \inf_{y \notin d_i} \max(\gamma_R(x, y), \mu_{d_i}(y)) = \inf_{y \notin d_i} \max(\gamma_R(x, y), 0) = \inf_{y \notin d_i} \gamma_R(x, y)$. Hence, $\mu_{\underline{R}(d_i)}(x) = \inf_{y \notin d_i}(\gamma_R(x, y))$.
2) It holds that $\gamma_{\underline{R}(d_i)}(x) = \sup_{y \in U} \min(\mu_R(x, y), \gamma_{d_i}(y))$. If $y \in d_i$, then $\gamma_{d_i}(y) = 0$, which leads to that $\min(\mu_R(x, y), \gamma_{d_i}(y)) = \min(\mu_R(x, y), 0) = 0$. We have that $\gamma_{\underline{R}(d_i)}(x) = \sup_{y \notin d_i} \min(\mu_R(x, y), \gamma_{d_i}(y)) = \sup_{y \notin d_i} \min(\mu_R(x, y), 1) = \sup_{y \notin d_i} \mu_R(x, y)$. This means that $\gamma_{\underline{R}(d_i)}(x) = \sup_{y \notin d_i} \mu_R(x, y)$.
3) With the fact that $y \notin d_i$ implies $\mu_{d_i}(y) = 0$ and $\min(\mu_R(x, y), \mu_{d_i}(y)) = 0$. Hence, $\mu_{\overline{R}(d_i)}(x) = \sup_{y \in U} \min(\mu_R(x, y), \mu_{d_i}(y)) = \sup_{y \in d_i} \min(\mu_R(x, y), \mu_{d_i}(y)) = \sup_{y \in d_i} \min(\mu_R(x, y), 1) = \sup_{y \in d_i} \mu_R(x, y)$. It implies that $\mu_{\overline{R}(d_i)}(x) = \sup_{y \in d_i} \mu_R(x, y)$.
4) With the fact that if $y \notin d_i$, then $\gamma_{d_i}(y) = 1$ and $\max(\mu_R(x, y), \gamma_{d_i}(y)) = 1$. Subsequently, $\gamma_{\overline{R}(d_i)}(x) = \inf_{y \in d_i} \max(\gamma_R(x, y), \gamma_{d_i}(y)) = \inf_{y \in d_i} \max(\gamma_R(x, y), 0) = \inf_{y \in d_i} \gamma_R(x, y)$. It follows that $\gamma_{\overline{R}(d_i)}(x) = \inf_{y \in d_i} \gamma_R(x, y)$. ■

Theorem 5 intuitively indicates the following.

1) The lower membership degree of an object $x$ to a decision class $d_i$ is the minimum diversity between $x$ and the objects outside the decision class.
2) The lower nonmembership degree of an object to a decision class is the maximal similarity between $x$ and the objects outside the decision class.
3) The upper membership degree of an object to a decision class is the maximal similarity between $x$ and the objects in the same decision class.
4) The upper nonmembership degree of an object to a decision class is the minimum diversity between $x$ and the objects in the same decision class. These conclusions present a simpler means of calculating the IF approximations of decision classes.

In particular, if an object does (or does not) belong to a decision class, the rough approximations are directly obtained as follows.

*Property 3:* Let $S = (U, \mathcal{R} \cup \{d\})$ be an IF decision system with $R \in \mathcal{R}$ and $U/d = \{d_1, d_2, \ldots, d_l\}$. For any $1 \le i \le l$, if $x \notin d_i$, then

$$\mu_{\underline{R}(d_i)}(x) = 0, \qquad \gamma_{\underline{R}(d_i)}(x) = 1.$$

*Proof:*

1) With the fact of $\gamma_R(x, x) = 0$ and $x \notin d_i$, it follows from Theorem 5 that $\mu_{\underline{R}(d_i)}(x) = \inf_{y \notin d_i}(\gamma_R(x, y)) = \gamma_R(x, x) = 0$.
2) With the fact of $\mu_R(x, x) = 0$ and $x \notin d_i$, we have that $\gamma_{\underline{R}(d_i)}(x) = \sup_{y \notin d_i}(\mu_R(x, y)) = \mu_R(x, x) = 1$. ■

Property 3 means that, if $x \notin d_i$, the lower membership degree and lower nonmembership degree are fixed to 0 and 1, respectively.

*Property 4:* Let $S = (U, \mathcal{R} \cup \{d\})$ be an IF decision system with $R \in \mathcal{R}$ and $U/d = \{d_1, d_2, \ldots, d_l\}$. For any $1 \le i \le l$, if $x \in d_i$, then

$$\mu_{\overline{R}(d_i)}(x) = 1, \qquad \gamma_{\overline{R}(d_i)}(x) = 0.$$

*Proof:*

1) We have that $\mu_R(x, x) = 0$. From Theorem 5, it holds that $\mu_{\overline{R}(d_i)}(x) = \sup_{y \in d_i}(\mu_R(x, y)) = \mu_R(x, x) = 1$.
2) Since $\gamma_R(x, x) = 0$, from Theorem 5, we see that $\gamma_{\overline{R}(d_i)}(x) = \inf_{y \in d_i}(\gamma_R(x, y)) = \gamma_R(x, x) = 0$. ■

Property 4 shows that if $x \in d_i$, the lower membership degree and lower nonmembership degree are fixed to 1 and 0, respectively.

From Theorem 5 and Properties 3 and 4, we find that the IF rough approximations in IF decision systems satisfy special properties, which will simplify the process of computation. In other words, if we want to obtain the rough approximations of an IF set, we need only to compute the lower approximations of an object w.r.t. its decision class, i.e., $\mu_{\underline{R}(d_i)}(x)$ and $\gamma_{\underline{R}(d_i)}(x)$ for $x \in d_i$, and the upper approximations of an object w.r.t. other decision classes, i.e., $\mu_{\overline{R}(d_i)}(x)$ and $\gamma_{\overline{R}(d_i)}(x)$ for $x \notin d_i$.

## IV. Knowledge Reduction of Intuitionistic Fuzzy Decision Systems

To measure the IF positive region generated by any subset $\mathcal{B} \subseteq \mathcal{R}$, define a dependence function $f(\mathcal{B}) = p(\text{Pos}_{\mathcal{B}}(d))$ as the probability quality of the IF positive region. The dependence function is the ratio of the sizes of the IF positive region. The reduct of IF decision systems can also be defined from the viewpoint of the dependence function as follows:

1) $f(\text{Pos}_{\mathcal{B}}(d)) = f(\text{Pos}_{\mathcal{R}}(d))$
2) $f(\text{Pos}_{\mathcal{B}-\{R\}}(d)) < f(\text{Pos}_{\mathcal{B}}(d)) \ \forall R \in \mathcal{B}$.

The following conclusions present a simplified method for computing the IF positive region and the dependence function.

*Property 5:* Let $S = (U, \mathcal{R} \cup \{d\})$ be an IF decision system with $U/d = \{d_1, d_2, \ldots, d_l\}$ and $\mathcal{B} \subseteq \mathcal{R}$. Then

1) $\text{Pos}_{\mathcal{B}}(d)(x) = \underline{\mathbf{B}}(d_{i_0})(x)$ where $x \in d_{i_0}$,
2) $f(\mathcal{B}) = \sum_{i=1}^{l} \sum_{x \in d_i} p(\underline{\mathbf{B}}(d_i)(x))$, and
3) $f(\mathcal{B}) = \frac{1}{2} + \frac{1}{2|U|} \sum_{i=1}^{l} \sum_{x \in d_i} (\mu_{\underline{\mathbf{B}}(d_i)}(x) - \gamma_{\underline{\mathbf{B}}(d_i)}(x))$.

*Proof:*

1) From Property 3, if $x \notin d_i$, then $\mu_{\underline{\mathbf{B}}(d_i)}(x) = 0$ and $\gamma_{\underline{\mathbf{B}}(d_i)}(x) = 1$. It follows that $\text{Pos}_{\mathcal{B}}(d)(x) = (\cup_{i=1}^{l} \underline{\mathbf{B}}(d_i))(x) = \underline{\mathbf{B}}(d_{i_0}))(x)$ for $x \in d_{i_0}$.
2) It holds that $f(\mathcal{B}) = P(\text{Pos}_{\mathcal{B}}(d)) = \sum_{x \in U} P(\text{Pos}_{\mathcal{B}}(d)(x))$. Combing with (1), we have that $f(\mathcal{B}) = \sum_{x \in U} P(\text{Pos}_{\mathcal{B}}(d)(x)) = \sum_{i=1}^{l} \sum_{x \in d_i} P(\underline{\mathbf{B}}(d_i))(x))$.
3) With the fact of $P(\underline{\mathbf{B}}(d_i)(x)) = \frac{1}{|U|} \frac{1 + \mu_{\underline{\mathbf{B}}(d_i)}(x) - \gamma_{\underline{\mathbf{B}}(d_i)}(x)}{2}$, it follows that

$$\begin{aligned} f(\mathcal{B}) &= \sum_{i=1}^{l} \sum_{x \in d_i} P(\underline{\mathbf{B}}(d_i)(x)) \\ &= \frac{1}{|U|} \sum_{i=1}^{l} \sum_{x \in d_i} \frac{1 + \mu_{\underline{\mathbf{B}}(d_i)}(x) - \gamma_{\underline{\mathbf{B}}(d_i)}(x)}{2} \\ &= \frac{1}{2} + \frac{1}{2|U|} \sum_{i=1}^{l} \sum_{x \in d_i} (\mu_{\underline{\mathbf{B}}(d_i)}(x) - \gamma_{\underline{\mathbf{B}}(d_i)}(x)). \blacksquare \end{aligned}$$

It is observed from Property 5 that if we want to obtain the IF positive region, we need only to compute the lower approximation of each object w.r.t. its own decision class.

The following example is employed to exhibit the idea of Property 5.

*Example 6:* Continued from Example 1.

Let $S = (U, \mathcal{R} \cup \{d\})$ be an IF decision system, where $U = \{x_1, x_2, x_3\}$ and $\mathcal{R} = \{R_i | 1 \leq i \leq 4\}$:

$$R_1 = \begin{pmatrix} (1,0) & (0.6,0.2) & (0.3,0.6) \\ (0.6,0.2) & (1,0) & (0.4,0.5) \\ (0.3,0.6) & (0.4,0.5) & (1,0) \end{pmatrix}$$

$$R_2 = \begin{pmatrix} (1,0) & (0.7,0.3) & (0.4,0.4) \\ (0.7,0.3) & (1,0) & (0.5,0.3) \\ (0.4,0.4) & (0.5,0.3) & (1,0) \end{pmatrix}$$

$$R_3 = \begin{pmatrix} (1,0) & (1,0) & (0.5,0.5) \\ (1,0) & (1,0) & (0.4,0.4) \\ (0.5,0.5) & (0.4,0.4) & (1,0) \end{pmatrix}$$

$$R_4 = \begin{pmatrix} (1,0) & (0.5,0.1) & (0.2,0.7) \\ (0.5,0.1) & (1,0) & (0.8,0.1) \\ (0.2,0.7) & (0.8,0.1) & (1,0) \end{pmatrix}.$$

Denote a special IF relation by $\mathbf{B} = \cap\{R | R \in \mathcal{B}\}$ w.r.t. any subset $\mathcal{B} \subseteq \mathcal{R}$. Clearly, $\mathbf{B}$ is a commonly used one in the existing literature. Consequently

$$\mathbf{R} = \cap\mathcal{R} = \begin{pmatrix} (1,0) & (0.5,0.3) & (0.2,0.7) \\ (0.5,0.3) & (1,0) & (0.4,0.5) \\ (0.2,0.7) & (0.4,0.5) & (1,0) \end{pmatrix}.$$

The decision is defined as $U/d = \{d_1, d_2\}$, where $d_1 = \{x_1, x_2\}$, $d_2 = \{x_3\}$.

From Property 3, we directly obtain that

$$\begin{aligned} &\mu_{\underline{R}(d_1)}(x_3) = 0, \quad \gamma_{\underline{R}(d_1)}(x_3) = 1 \\ &\mu_{\underline{R}(d_2)}(x_1) = 0, \quad \gamma_{\underline{R}(d_2)}(x_1) = 1 \\ &\mu_{\underline{R}(d_2)}(x_2) = 0, \quad \gamma_{\underline{R}(d_2)}(x_2) = 1. \end{aligned}$$

From Theorem 5, we compute that

$$\begin{aligned} &\mu_{\underline{\mathbf{R}}(d_1)}(x_1) = \inf_{x_i \notin d_1} \gamma_{\mathbf{R}}(x_1, x_i) \\ &= \inf(\gamma_{\mathbf{R}}(x_1, x_2), \gamma_{\mathbf{R}}(x_1, x_3)) = \inf(0.3, 0.7) = 0.3 \\ &\gamma_{\underline{\mathbf{R}}(d_1)}(x_1) = \sup_{x_i \notin d_1} (\mu_{\mathbf{R}}(x_1, x_i)) \\ &= \sup(\mu_{\mathbf{R}}(x_1, x_2), \mu_{\mathbf{R}}(x_1, x_3)) = \sup(0.5, 0.2) = 0.5 \\ &\mu_{\underline{\mathbf{R}}(d_2)}(x_2) = \inf_{x_i \notin d_2} \gamma_{\mathbf{R}}(x_2, x_i) = \inf(\gamma_{\mathbf{R}}(x_2, x_1)) \\ &= \inf(0.3) = 0.3 \\ &\gamma_{\underline{\mathbf{R}}(d_2)}(x_2) = \sup_{x_i \notin d_2} (\mu_{\mathbf{R}}(x_2, x_i)) = \sup(\mu_{\mathbf{R}}(x_2, x_1)) \\ &= \sup(0.5) = 0.5 \\ &\mu_{\underline{\mathbf{R}}(d_2)}(x_3) = \inf_{x_i \notin d_2} \gamma_{\mathbf{R}}(x_3, x_i) = \inf(\gamma_{\mathbf{R}}(x_3, x_1)) \\ &= \inf(0.7) = 0.7 \\ &\gamma_{\underline{\mathbf{R}}(d_2)}(x_3) = \sup_{x_i \notin d_2} (\mu_{\mathbf{R}}(x_3, x_i)) = \sup(\mu_{\mathbf{R}}(x_3, x_1)) \\ &= \sup(0.2) = 0.2. \end{aligned}$$

To sum up

$$\begin{aligned} \underline{\mathbf{R}}(d_1) &= \frac{(0.3, 0.5)}{x_1} + \frac{(0, 1)}{x_2} + \frac{(0, 1)}{x_3} \\ \underline{\mathbf{R}}(d_2) &= \frac{(0, 1)}{x_1} + \frac{(0.3, 0.5)}{x_2} + \frac{(0.7, 0.2)}{x_3}. \end{aligned}$$

The IF positive region w.r.t. $\mathcal{R}$ is

$$\mathrm{Pos}_{\mathcal{R}}(d) = \underline{\mathbf{R}}(d_1) \cup \underline{\mathbf{R}}(d_2) = \frac{(0.3, 0.5)}{x_1} + \frac{(0.3, 0.5)}{x_2} + \frac{(0.7, 0.2)}{x_3}.$$

The probability quality of the IF positive region is

$$p(\mathrm{Pos}_{\mathcal{R}}(d)) = \frac{|\mathrm{Pos}_{\mathcal{R}}(d)|}{|U|} = \frac{\sum_{x \in U} \frac{1+\mu_{\mathrm{Pos}_{\mathcal{R}}(d)}(x) - \gamma_{\mathrm{Pos}_{\mathcal{R}}(d)}(x)}{2}}{|U|}$$
$$= \frac{1}{3}\left(\frac{1+0.3-0.5}{2} + \frac{1+0.3-0.5}{2} + \frac{1+0.7-0.2}{2}\right) = 0.51.$$

*Theorem 6:* Let $S = (U, \mathcal{R} \cup \{d\})$ be an IF decision system with $\mathcal{A}$, $\mathcal{B} \subseteq \mathcal{R}$. If $\mathbf{B} \preceq \mathbf{A}$, then

1) $\mathrm{Pos}_{\mathcal{A}}(d) \subseteq \mathrm{Pos}_{\mathcal{B}}(d)$;
2) $f(\mathcal{A}) \leq f(\mathcal{B})$.

*Proof:*

1) Since $\mathbf{B} \preceq \mathbf{A}$, it holds from Definition 3 that $\underline{\mathbf{A}}(d_i)(x) \subseteq \underline{\mathbf{B}}(d_i)(x)$ for any $d_i$. This implies that $\mathrm{Pos}_{\mathcal{A}}(d) \subseteq \mathrm{Pos}_{\mathcal{B}}(d)$.
2) It can be concluded from (1). ■

It is noted from Theorem 6 that the IF positive region is monotonic with the granularity of the induced IF relation. In other words, the finer is the IF relation induced by a subset of IF relations, the larger is the IF positive region. A subset of IF relations may maintain the entire positive region. From this viewpoint, the notion of reduct can be introduced.

The dependence function reflects the classification ability of IF relations and can be used to evaluate the significance of IF relations. Adding different IF relations to a given subset of IF relations may lead to a variety of different changes in the dependence function. Accordingly, the significance of IF relations can be defined as follows.

*Definition 8:* Let $S = (U, \mathcal{R} \cup \{d\})$ be an IF decision system with $\mathcal{B} \subseteq \mathcal{R}$ and $R \in \mathcal{R} - \mathcal{B}$. The significance of $R$ w.r.t. $\mathcal{B}$ is defined as $\mathrm{Sig}(R, \mathcal{B}, d) = f(\mathcal{B} \cup \{R\}) - f(\mathcal{B})$.

Based on the above-mentioned discussion, a forward heuristic algorithm for finding a reduct of an IF decision system is formulated as shown in Algorithm 1.

**Algorithm 1** A Heuristic Algorithm for Finding a Reduct of an IF Decision System Based on IF Positive Region (Algorithm IFPR).

Input: An IF decision system $(U, \mathcal{R} \cup \{d\})$;
Output: An IF relation reduct $\mathcal{B}$.
1: Let $\mathcal{B} = \emptyset$;
2: **For** each IF relation $R \in \mathcal{R}$
3: Compute the IF relation $\mathcal{B} \cup R$;
4: Compute the IF approximations $\mu_{\underline{\mathcal{B} \cup R}(d_i)}(x)$ and $\gamma_{\underline{\mathcal{B} \cup R}(d_i)}(x)$ for each $x \in d_i$ according to Theorem 5;
5: Compute the dependence functions $f(\mathcal{B})$ and $f(\mathcal{B} \cup \{R\})$ according to Property 5(3);
6: Compute the significance $\mathrm{Sig}(R, \mathcal{B}, d)$ according to Definition 8;
7: End **For**
8: Find $R_0$ with maximum value $\mathrm{Sig}(R_0, \mathcal{B}, d)$;
9: If $\mathrm{Sig}(R_0, \mathcal{B}, d) > \delta$;
10: Let $\mathcal{B} \leftarrow \mathcal{B} \cup \{R_0\}$ and $\mathcal{R} \leftarrow \mathcal{R} - \{R_0\}$;
11: Return to Step 2;
12: Else
13: Output $\mathcal{B}$ and terminate the algorithm.
14: End If

The parameter $\delta$ is used to stop the loop in the algorithm. The larger is the value of $\delta$, the larger is the number of selected IF relations. Therefore, the value of $\delta$ should be chosen in appropriate ranges according to real data. Moreover, the computation of IF relation $\mathcal{B} \cup R$ in Step 3 adopts an incremental method. That is, we remember the IF relation $\mathcal{B}$ in the previous loop and take an intersection between it and relation $R$, instead of recalculating the IF relation $\mathcal{B} \cup R$. This strategy avoids a considerable amount of unnecessary time-consuming calculation.

In Step 4, the computation of the IF approximations for every object can be completed in $O(|U|^2)$. In Step 5, the dependence functions can be obtained within $O(|U|^2)$. In Steps 9–11, the elements of $\mathcal{R}$ are evaluated by using the significance measure, and their complexity is $O(|U|^2|\mathcal{R}|)$. The overall time complexity of the algorithm is thus $O(|U|^2|\mathcal{R}|)$.

Let us employ an example to show the idea of the new algorithm.

*Example 7:* Continued from Example 6.

We take the value $\delta = 0$ in Algorithm 1 for illustration.

We first initialize $\mathcal{B} = \emptyset$, and compute the IF approximations under each $R_i (1 \leq i \leq 4)$, respectively, in what follows.

We obtain that

$$\mu_{\underline{R_1}(d_1)}(x_1) = \inf_{x_i \notin d_1} \gamma_{R_1}(x_1, x_i)$$
$$= \inf(\gamma_{R_1}(x_1, x_2), \gamma_{R_1}(x_1, x_3)) = \inf(0.2, 0.6) = 0.2$$
$$\gamma_{\underline{R_1}(d_1)}(x_1) = \sup_{x_i \notin d_1} (\mu_{R_1}(x_1, x_i))$$
$$= \sup(\mu_{R_1}(x_1, x_2), \mu_{R_1}(x_1, x_3)) = \sup(0.6, 0.3) = 0.6$$
$$\mu_{\underline{R_1}(d_2)}(x_2) = \inf_{x_i \notin d_2} \gamma_{R_1}(x_2, x_i)$$
$$= \inf(\gamma_{R_1}(x_2, x_1)) = \inf(0.3) = 0.3$$
$$\gamma_{\underline{R_1}(d_2)}(x_2) = \sup_{x_i \notin d_2} (\mu_{R_1}(x_2, x_i))$$
$$= \sup(\mu_{R_1}(x_2, x_1)) = \sup(0.6) = 0.6$$
$$\mu_{\underline{R_1}(d_2)}(x_3) = \inf_{x_i \notin d_2} \gamma_{R_1}(x_3, x_i)$$
$$= \inf(\gamma_{R_1}(x_3, x_1)) = \inf(0.6) = 0.6$$
$$\gamma_{\underline{R_1}(d_2)}(x_3) = \sup_{x_i \notin d_2} (\mu_{R_1}(x_3, x_i))$$
$$= \sup(\mu_{R_1}(x_3, x_1)) = \sup(0.3) = 0.3.$$

Hence

$$\underline{R_1}(d_1) = \frac{(0.2, 0.6)}{x_1} + \frac{(0, 1)}{x_2} + \frac{(0, 1)}{x_3}$$
$$\underline{R_1}(d_2) = \frac{(0, 1)}{x_1} + \frac{(0.3, 0.6)}{x_2} + \frac{(0.6, 0.3)}{x_3}.$$

The IF positive region w.r.t. $R_1$ is

$$\begin{aligned}\mathrm{Pos}_{R_1}(d) &= \underline{R_1}(d_1) \cup \underline{R_1}(d_2)\\ &= \frac{(0.2, 0.6)}{x_1} + \frac{(0.3, 0.6)}{x_2} + \frac{(0.6, 0.3)}{x_3}.\end{aligned}$$

The dependence function of the IF positive region w.r.t. $R_1$ is

$$\begin{aligned}f(R_1) = p(\mathrm{Pos}_{R_1}(d)) &= \frac{|\mathrm{Pos}_{R_1}(d)|}{|U|}\\ &= \frac{\sum_{x\in U}\frac{1+\mu_{\mathrm{Pos}_{R_1}(d)}(x)-\gamma_{\mathrm{Pos}_{R_1}(d)}(x)}{2}}{|U|} = \frac{1}{3}\bigg(\frac{1+0.2-0.6}{2}\\ &\quad + \frac{1+0.3-0.6}{2} + \frac{1+0.6-0.3}{2}\bigg) = 0.33.\end{aligned}$$

Similarly

$$\underline{R_2}(d_1) = \frac{(0.3, 0.7)}{x_1} + \frac{(0, 1)}{x_2} + \frac{(0, 1)}{x_3}$$
$$\underline{R_2}(d_2) = \frac{(0, 1)}{x_1} + \frac{(0.3, 0.7)}{x_2} + \frac{(0.4, 0.4)}{x_3}.$$

The IF positive region w.r.t. $R_2$ is

$$\begin{aligned}\mathrm{Pos}_{R_2}(d) &= \underline{R_2}(d_1) \cup \underline{R_2}(d_2)\\ &= \frac{(0.3, 0.7)}{x_1} + \frac{(0.3, 0.7)}{x_2} + \frac{(0.4, 0.4)}{x_3}.\end{aligned}$$

The dependence function of the IF positive region w.r.t. $R_2$ is $f(R_2) = p(\mathrm{Pos}_{R_2}(d)) = 0.37$.

In addition

$$\underline{R_3}(d_1) = \frac{(0, 1)}{x_1} + \frac{(0, 1)}{x_2} + \frac{(0, 1)}{x_3}$$
$$\underline{R_3}(d_2) = \frac{(0, 1)}{x_1} + \frac{(0, 1)}{x_2} + \frac{(0.5, 0.5)}{x_3}.$$

The IF positive region w.r.t. $R_3$ is

$$\mathrm{Pos}_{R_1}(d) = \underline{R_3}(d_1) \cup \underline{R_3}(d_2) = \frac{(0, 1)}{x_1} + \frac{(0, 1)}{x_2} + \frac{(0.5, 0.5)}{x_3}.$$

The dependence function of the IF positive region w.r.t. $R_3$ is $f(R_3) = p(\mathrm{Pos}_{R_2}(d)) = 0.17$.

Additionally

$$\underline{R_4}(d_1) = \frac{(0.1, 0.5)}{x_1} + \frac{(0, 1)}{x_2} + \frac{(0, 1)}{x_3}$$
$$\underline{R_4}(d_2) = \frac{(0, 1)}{x_1} + \frac{(0.1, 0.5)}{x_2} + \frac{(0.7, 0.2)}{x_3}.$$

The IF positive region w.r.t. $R_4$ is

$$\mathrm{Pos}_{R_4}(d) = \frac{(0.1, 0.5)}{x_1} + \frac{(0.1, 0.5)}{x_2} + \frac{(0.7, 0.2)}{x_3}.$$

The dependence function of the IF positive region w.r.t. $R_4$ is $f(R_4) = p(\mathrm{Pos}_{R_4}(d)) = 0.45$.

By the above analyses, we further obtain that

$$\mathrm{Sig}(R_1, \mathcal{B}, d) = 0.33, \quad \mathrm{Sig}(R_2, \mathcal{B}, d) = 0.37$$
$$\mathrm{Sig}(R_3, \mathcal{B}, d) = 0.17, \quad \mathrm{Sig}(R_4, \mathcal{B}, d) = 0.45.$$

From Steps 8–10, we should set $\mathcal{B} = \{R_4\}$ and $\mathcal{R} = \{R_1, R_2, R_3\}$. Since $\mathrm{Sig}(R_4, \mathcal{B}, d) > 0$, we go to the next cycle.

It is calculated that

$$\mathrm{Pos}_{\mathcal{B}\cup R_1}(d) = \frac{(0.1, 0.5)}{x_1} + \frac{(0.1, 0.5)}{x_2} + \frac{(0.7, 0.2)}{x_3}$$
$$\mathrm{Pos}_{\mathcal{B}\cup R_2}(d) = \frac{(0.3, 0.5)}{x_1} + \frac{(0.3, 0.5)}{x_2} + \frac{(0.7, 0.2)}{x_3}$$
$$\mathrm{Pos}_{\mathcal{B}\cup R_3}(d) = \frac{(0.1, 0.5)}{x_1} + \frac{(0.1, 0.5)}{x_2} + \frac{(0.7, 0.2)}{x_3}.$$

Consequently

$$\mathrm{Sig}(R_1, \mathcal{B}, d) = 0.45 - 0.45 = 0$$
$$\mathrm{Sig}(R_2, \mathcal{B}, d) = 0.51 - 0.45 = 0.06$$
$$\mathrm{Sig}(R_3, \mathcal{B}, d) = 0.45 - 0.45 = 0.$$

Hence, $\mathcal{B} = \{R_4, R_2\}$ and $\mathcal{R} = \{R_1, R_3\}$. Since $\mathrm{Sig}(R_2, \mathcal{B}, d) > 0$, we next go to the third cycle.

With the fact that $\mathrm{Sig}(R_1, \mathcal{B}, d) = 0$ and $\mathrm{Sig}(R_3, \mathcal{B}, d) = 0$, the algorithm is terminated and an IF reduct $\mathcal{B} = \{R_4, R_2\}$ is returned.

## V. Numerical Experiments

In a sequence of experiments, we assess the performance of the proposed algorithm, IFPR, and compare it with some existing attribute reduction algorithms. The experiments are set up as follows.

1) The experiments are performed on a personal computer with Windows XP and an Intel (R) Core (TM) i5-2400 CPU @ 3.10 GHz with 4 GB of memory. The algorithms are implemented using MATLAB 7.8. Datasets downloaded from UCI machine learning repository [60] and Keng Ridge Bio-medical dataset repository [61] are used for experimental analysis, which are described in Table I.
2) Before reduction, datasets are transformed into IF decision systems. For convenience, we adopt a simple method by computing the similarity degree of objects. For any attribute $a$, the similarity degree between objects $x$ and $y$ is defined as [62]

$$R_a(x, y) = \max\left(\min\left(\frac{a(y) - a(x) + \sigma}{\sigma}, \frac{a(x) - a(y) + \sigma}{\sigma}\right), 0\right)$$

where $\sigma^2$ is the variance of attribute a.

For any attribute set $B$, the corresponding IF relation $R_B$ is defined as $R_B = \{< (x, y), \mu_{R_B}(x, y), \gamma_{R_B}(x, y) > |(x, y) \in U \times U\}$, where $\mu_{R_B}(x, y) = \min_{a\in B} R_a(x, y)$ and $\gamma_{R_B}(x, y) = 1 - \max_{a\in B} R_a(x, y)$. Obviously, it

TABLE I
DESCRIPTION OF DATA

| No. | Data sets | Samples | Features | Decision classes |
|---|---|---|---|---|
| 1 | Wine | 178 | 13 | 3 |
| 2 | heart | 270 | 13 | 2 |
| 3 | hepatitis | 155 | 19 | 2 |
| 4 | ICU | 200 | 20 | 3 |
| 5 | wpbc | 194 | 33 | 2 |
| 6 | australian_credit | 690 | 15 | 9 |
| 7 | Soner | 208 | 60 | 2 |
| 8 | horse | 368 | 22 | 2 |
| 9 | wdbc | 569 | 30 | 2 |
| 10 | iono | 351 | 34 | 2 |
| 11 | autovalve_B | 414 | 72 | 3 |
| 12 | Colon | 62 | 2000 | 2 |
| 13 | Breast | 84 | 9216 | 5 |
| 14 | Leukemia | 72 | 7129 | 2 |
| 15 | MLL | 72 | 12582 | 3 |

TABLE II
COMPARISON ON THE NUMBERS OF SELECTED ATTRIBUTES

| Data sets | IFPR | B-FRFS | DM-FR | FA-FPR | Raw data |
|---|---|---|---|---|---|
| Wine | **6** | 11 | 12 | 11 | 13 |
| heart | **8** | 11 | 12 | 11 | 13 |
| hepatitis | **8** | 9 | 13 | 9 | 19 |
| ICU | 11 | 10 | **9** | 11 | 20 |
| wpbc | **10** | 22 | 26 | 19 | 33 |
| australian_credit | **8** | 10 | 10 | 10 | 15 |
| Soner | **13** | 20 | 35 | 17 | 60 |
| horse | **9** | 13 | 16 | 10 | 23 |
| wdbc | **17** | 24 | 23 | 16 | 10 |

23

wdbc **17**

TABLE IV
COMPARISON ON CLASSIFICATION ACCURACIES OF REDUCED DATA WITH CART

| Data sets | IFPR | B-FRFS | DM-FR | FA-FPR | Raw data |
|---|---|---|---|---|---|
| Wine | 92.08±4.81 | 91.53±6.09 | 90.97±6.05 | **92.42±6.50** | 89.86±6.35 |
| heart | **74.81±8.15** | 73.70±6.16 | 74.44±6.16 | 74.07±6.05 | 74.07±6.30 |
| hepatitis | **91.67±6.14** | 89.67±5.57 | 89.67±3.33 | 91.00±4.46 | 84.83±8.69 |
| ICU | **82.45±3.09** | 81.45±3.09 | 81.92±3.11 | **82.45±3.09** | 79.40±3.16 |
| wpbc | **74.24±7.66** | 68.18±10.37 | 71.63±9.22 | 70.68±6.29 | 69.63±6.78 |
| australian_credit | **83.17±4.24** | 82.45±5.59 | 82.45±4.79 | 81.73±5.72 | 82.45±4.79 |
| Soner | 68.71±9.55 | 72.60±11.68 | **73.62±11.61** | 72.24±12.29 | 73.02±14.91 |
| horse | 95.93±1.90 | **96.19±1.94** | 90.47±4.35 | 88.33±3.96 | 96.19± 2.28 |
| wdbc | 92.96±3.12 | 92.10±2.20 | 92.08±3.45 | **93.84±2.09** | 90.50±4.55 |
| iono | **90.96±3.92** | 87.08±6.93 | 90.37±5.89 | 86.45±7.18 | 86.43±7.22 |
| autovalve_B | **86.49±6.73** | 86.16±4.95 | 84.89±6.68 | 84.24±6.22 | 85.83±5.41 |
| Colon | **90.00±14.05** | 88.33±15.81 | 78.75±16.01 | 81.25±15.12 | 80.00±21.94 |
| Breast | **84.58±11.46** | 80.43±10.43 | 82.25±10.94 | 77.50±16.46 | 63.75±12.43 |
| Leukemia | 75.56±16.53 | 77.19±9.19 | **77.97±23.05** | **77.97±23.05** | 75.56±16.53 |
| MLL | **97.14±6.02** | 92.71±10.44 | 91.21±11.03 | 90.43±15.06 | 89.86±15.33 |
| Average | **85.38**±7.07 | 83.99±7.36 | 83.51±8.38 | 82.97±8.9 | 81.43±9.20 |

TABLE V
COMPARISON ON CLASSIFICATION ACCURACIES OF REDUCED DATA WITH KNN

| Data sets | IFPR | B-FRFS | DM-FR | FA-FPR | Raw data |
|---|---|---|---|---|---|
| Wine | 97.22±2.93 | 96.04±2.74 | 95.49±4.41 | **97.71±2.97** | 95.49±3.54 |
| heart | **85.19±6.30** | 81.85±6.40 | 83.33±6.36 | 74.44±6.16 | 82.59±6.06 |
| hepatitis | **86.50±6.31** | 85.00±6.63 | 81.67±7.17 | 82.83±5.56 | 80.40±5.84 |
| ICU | **93.08±2.31** | 91.21±2.25 | 91.21±2.25 | 91.21±2.25 | 88.61±1.25 |
| wpbc | **76.63±8.23** | 73.68±6.87 | 75.79±5.58 | 74.71±7.95 | 76.26±5.89 |
| australian_credit | **81.30±5.02** | 78.27±2.27 | 79.27±3.18 | 80.23±5.82 | 75.25±4.79 |
| Soner | **75.98±5.43** | 73.57±5.57 | 72.14±6.94 | 74.07±8.94 | 72.62±7.05 |
| horse | 89.13±3.61 | **91.02±3.67** | 86.45±5.86 | 89.68±5.05 | 88.05±5.58 |
| wdbc | 96.15±3.04 | 96.50±2.18 | 96.50±2.46 | **96.67±3.00** | 93.33±2.09 |
| iono | **84.12±3.92** | 82.13±5.56 | 82.54±4.70 | 82.72±5.70 | 82.40±5.02 |
| autovalue_B | 91.31±4.57 | 90.66±4.67 | 91.35±4.09 | **91.95±5.09** | 90.98±4.51 |
| Colon | 84.58±18.84 | **87.08±17.17** | 76.25±12.12 | 79.58±11.19 | 73.33±14.05 |
| Breast | 78.33±11.08 | 86.67±13.86 | **90.42±8.84** | 82.50±10.54 | 66.25±10.29 |
| Leukemia | 80.42±6.98 | 76.19±6.62 | **84.22±6.46** | 83.31±8.54 | 80.42±6.98 |
| MLL | **98.57±4.52** | 92.29±10.69 | 96.73±8.14 | 97.14±6.02 | 79.95±15.97 |
| Average | **86.57**±6.21 | 85.54±6.48 | 85.55±5.90 | 85.25±6.32 | 81.73±6.59 |

the dimensionality reduction achieves $50\%$ for low-dimensional datasets and $99.98\%$ for high-dimensional datasets. This implies that all the reduction algorithms are efficient in terms of reducing redundant attributes. Moreover, as seen in Fig. 1, the results obtained by the proposed algorithm are consistently better than those obtained by the other algorithms on most datasets. For example, the proposed algorithm selects eight attributes on dataset heart and ten attributes on dataset wpbc, while the other algorithms select more than ten attributes on most data. The number of attributes selected on an average by the new algorithm is 11.47, which is smaller than that selected by the other three algorithms. Hence, the proposed algorithm is effective in terms of decreasing the dimensionality of data.

Table III records the computational time of the reduction algorithms, which is depicted in Fig. 2. We can see that algorithm IFPR and algorithm FA-FPR are faster than the other two in most cases. This is because that algorithm FA-FPR is equipped with an accelerator, where the numbers of samples that need to be dealt with in the current step is largely reduced in previous steps, whereas the proposed algorithm adopts a fast computation method based on matrix representation and matrix logic. As seen in this table, algorithm IFPR takes dozens of seconds on all datasets. For instance, algorithm IFPR takes about 32.16 s on data Leukemia, 0.39 s on data iono, and 27.14 s on data MLL. These results indicate that algorithm IFPR runs fast on both low-dimensional and high-dimensional datasets. Hence, the proposed algorithm is more acceptable in terms of running time.

Tables IV and V record the classification performance of the four reduction algorithms tested using two different types of classifiers CART and KNN ($K = 10$), respectively, where the black-bordered symbols indicate the highest among the obtained classification accuracies. In the case of CART, all the four algorithms significantly improve the classification accuracies on most datasets as compared to the raw data. Moreover, algorithm IFPR outperforms the other algorithms for nine times among all the 15 datasets. In the same time, algorithm IFPR outperforms the other algorithms for eight times with respect to KNN. It should be emphasized that algorithm IFPR shows a decrease in the classification accuracy as compared to the raw horse dataset in Table IV. However, the decrease in the classification accuracy is only 0.26%, while that achieves in dimensionality 60.87%. On the other 14 datasets, algorithm IFPR offers a clear improvement in the classification accuracy as compared to the raw data. Moreover, the average accuracy indicates that algorithm IFPR ranks first with a margin of 1.39% between it and the second best accuracy 83.99% of algorithm B-FRFS, and a margin of 3.94% between it and the 81.43% accuracy rate of the raw data with

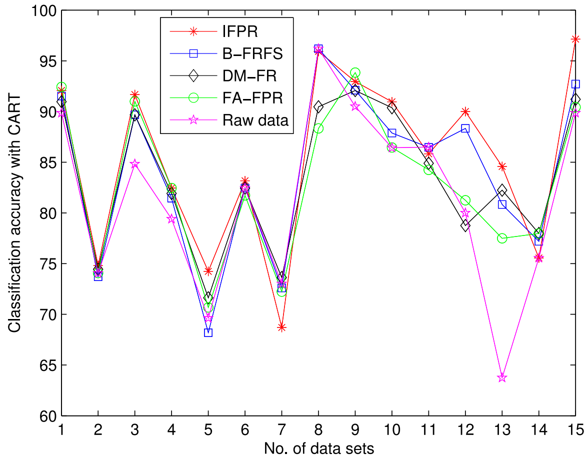


Fig. 3. Classification accuracy of reduced data with CART.

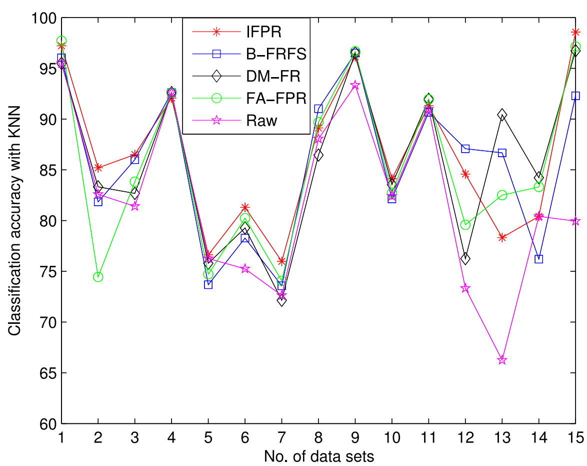


Fig. 4. Classification accuracy of reduced data with KNN.

respect to CART. Meanwhile, the average accuracy achieved by algorithm IFPR is 86.57%, which exceeds that obtained by algorithm DM-FR by a margin of 1.02% and the raw data by a margin of $4.84\%$ with respect to KNN. Figs. 3 and 4 display the classification accuracies of the four algorithms and the raw data based on the CART and KNN classifiers, respectively. It can be seen that IFPR gains favorable classification results on most of the datasets. Thus, the proposed method, which takes the similarity and diversity of objects into account, is feasible and effective. The reasons can be elaborated from the following aspects. Algorithm B-FRFS strives to achieve the smallest uncertainty of the decision classes. This idea can decrease the difference of the lower and upper approximations of all decision concepts. Algorithm DM-FR uses the index of the occurrence frequency of features in the discernibility matrix. The subset of features selected depends on the discernibility matrix constructed. Algorithm FA-FPR is aimed to maintain the positive region unchanged, which can guarantee the maximal degree of samples' membership to their own classes. However, these algorithms are based on the same fuzzy rough set model, which considers only the relevance of samples while overlooking the diversity. The proposed algorithm is based on the IF rough set that takes both the two factors into account. This idea can maintain both the maximal degrees of samples' membership to their own classes and nonmembership to other classes, and thus can guarantee the maximum rate of correct classification and the minimum rate of error classification simultaneously.

## VI. Conclusion

The rough set and IF set are two important tools for describing the uncertainty of knowledge and have been widely applied in the frameworks of granular computing and data dimensionality reduction. In this study, we introduced several types of fuzzy granules, utilizing which granular structures of the lower and upper approximations of the IF set were addressed. Furthermore, to deal with data dimensionality reduction with IF rough set, we formalized a type of information systems, i.e., IF decision information systems, for representing IF information. The IF approximations were examined in IF decision systems. The knowledge reduction of IF decision systems was addressed from the viewpoint of maintaining the IF approximations of decision classes, and a reduction algorithm was developed to reduce redundant IF relations in these systems. To examine the efficiency of the proposed algorithm, we constructed IF relations from numerical attributes of data, which bear both the similarity and diversity degrees between objects. Experiments were conducted to compare the new algorithm with existing ones. This study presented a pioneering attempt to investigate the problem of attribute reduction on real datasets by combining the techniques of the rough set and IF set.

Some problems related to the IF rough set still remain that need to be considered and discussed further. They are summarized as follows. First, the granular structures of the fuzzy rough set were elaborated for attribute reduction in many studies in the literature. How can we demonstrate the corresponding solution of granular structures of the IF rough set? Second, many IF relations can be induced from real datasets using substantial methods. Convergency is important for the theoretical analysis and algorithmic application of IF rough set models. In a future study, we will search the models that can maintain the convergency of algorithms on real datasets. Third, IF rough set-based techniques also need to be applied to the fields of classification learning in the future.

## References

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

[2] J. Dai, Q. Hu, J. Zhang, H. Hu, and N. Zheng, "Attribute selection for partially labeled categorical data by rough set approach," *IEEE Trans. Cybern.*, vol. 47, no. 9, pp. 2460–2471, Sep. 2017.

[3] J. Dai, Q. Hu, H. Hu, and D. Huang, "Neighbor inconsistent pair selection for attribute reduction by rough set approach," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 937–950, Apr. 2018.

[4] A. Tan, W. Wu, J. Li, and G. Lin, "Evidence-theory-based numerical characterization of multigranulation rough sets in incomplete information systems," *Fuzzy Sets Syst.*, vol. 294, no. 1, pp. 18–35, 2016.

[5] A. Tan, W. Wu, and Y. Tao, "A set cover-based approach for the test-cost-sensitive attribute reduction problem," *Soft Comput.*, vol. 21, no. 20, pp. 6159–6173, 2017.

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

[7] D. Dubois and H. Prade, "Fuzzy sets in approximate reasoning," *Fuzzy Sets Syst.*, vol. 40, no. 1, pp. 143–244, 1991.

[8] Q. Hu, D. Yu, W. Pedrycz, and D. Chen, "Kernelized fuzzy rough sets and their applications," *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 11, pp. 1649–1667, Nov. 2011.

[9] Q. Hu, L. Zhang, Y. Zhou, and W. Pedrycz, "Large-scale multimodality attribute reduction with multi-kernel fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 1, pp. 226–238, Feb. 2018.
[10] R. Jensen and Q. Shen, "Fuzzy-rough attributes reduction with application to web categorization," *Fuzzy Sets Syst.*, vol. 141, no. 3, pp. 469–485, 2004.
[11] R. Wang, D. Chen, and S. Wong, "Fuzzy-rough-set-based active learning," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 6, pp. 1699–1704, Dec. 2014.
[12] J. Mi, Y. Leung, H. Zhao, and T. Feng, "Generalized fuzzy rough sets determined by a triangular norm," *Inf. Sci.*, vol. 178, no. 16, pp. 3203–3213, 2008.
[13] W. Wu, Y. Leung, and J. Mi, "On characterizations of (I,T)-fuzzy rough approximation operators," *Fuzzy Sets Syst.*, vol. 154, no. 1, pp. 76–102, 2005.
[14] Y. Y. Yao, "A comparative study of fuzzy sets and rough sets," *Inf. Sci.*, vol. 109, no. 1–4, pp. 21–47, 1998.
[15] D. S. Yeung, D. Chen, E. C. C. Tsang, J. W. T. Lee, and X. Wang, "On the generalization of fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 13, no. 3, pp. 343–361, Jun. 2005.
[16] L. A. Zadeh, "Fuzzy sets," *Inf. Control*, vol. 8, no. 3, pp. 338–353, 1965.
[17] X. Liu, W. Pedrycz, T. Chai, and M. Song, "The development of fuzzy rough sets with the use of structures and algebras of axiomatic fuzzy sets," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 3, pp. 443–462, Mar. 2009.
[18] Y. Qian, Y. Li, J. Liang, G. Lin, and C. Dang, "Fuzzy granular structure distance," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 2245–2259, Dec. 2015.
[19] W. Xu and W. Li, "Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets," *IEEE Trans. Cybern.*, vol. 46, no. 2, pp. 366–379, Feb. 2016.
[20] D. Chen, Y. Yang, and H. Wang, "Granular computing based on fuzzy similarity relations," *Soft Comput.*, vol. 15, no. 6, pp. 1161–1172, 2010.
[21] Q. He, C. Wu, D. Chen, and S. Zhao, "Fuzzy rough set based attribute reduction for information systems with fuzzy decisions," *Knowl.-Based Syst.*, vol. 24, no. 5, pp. 689–696, 2011.
[22] Y. Lin, Q. Hu, J. Liu, J. Li, and X. Wu, "Streaming feature selection for multi-label learning based on fuzzy mutual information," *IEEE Trans. Fuzzy Syst.*, vol. 25, no. 6, pp. 1491–1507, Dec. 2017.
[23] R. Jensen and Q. Shen, "New approaches to fuzzy-rough feature selection," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 824–838, Aug. 2009.
[24] R. B. Bhatt and M. Gopal, "On fuzzy rough sets approach to feature selection," *Pattern Recognit. Lett.*, vol. 26, no. 7, pp. 965–975, 2005.
[25] G. C. Y. Tsang, D. Chen, E. C. C. Tsang, J. W. T. Lee, and D. S. Yeung, "On attributes reduction with fuzzy rough sets," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Oct. 2005, vol. 3, no. 3, pp. 2775–2780.
[26] D. Chen and 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, Oct. 2014.
[27] C. C. E. Tsang, D. Chen, S. D. Yueng, W. T. J. Lee, and X. Wang, "Attribute reduction using fuzzy rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 16, no. 5, pp. 1130–1141, Oct. 2008.
[28] C. Wang, M. Shao, Q. He, Y. Qian, and Y. Qi, "Feature subset selection based on fuzzy neighborhood rough sets," *Knowl.-Based Syst.*, vol. 111, no. 1, pp. 173–179, 2016.
[29] 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.
[30] C. Wang, X. Wang, D. Chen, Q. Hu, Z. Dong, and Y. Qian, "Feature selection based on neighborhood discrimination index," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 7, pp. 2986–2999, Jul. 2018.
[31] Q. Hu, L. Zhang, S. An, D. Zhang, and D. Yu, "On robust fuzzy rough set models," *IEEE Trans. Fuzzy Syst.*, vol. 20, no. 4, pp. 636–651, Aug. 2012.
[32] S. Zhao, H. Chen, C. Li, X. Du, and H. Sun, "A novel approach to building a robust fuzzy rough classifier," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 4, pp. 769–786, Aug. 2015.
[33] S. Zhao, E. C. C. Tsang, and D. Chen, "The model of fuzzy variable precision rough sets," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 2, pp. 451–467, Apr. 2009.
[34] Q. Hu, Z. Xie, and D. Yu, "Hybrid attribute reduction based on a novel fuzzy-rough model and information granulation," *Pattern Recognit.*, vol. 40, no. 12, pp. 3509–3521, 2007.
[35] X. Zhang, C. Mei, D. Chen, and J. Li, "Feature selection in mixed data: A method using a novel fuzzy rough set-based information entropy," *Pattern Recognit.*, vol. 56, no. 1, pp. 1–15, 2016.
[36] K. Atanassove, "Intuitionistic fuzzy sets," *Fuzzy Sets Syst.*, vol. 20, no. 1, pp. 87–96, 1986.
[37] K. Atanassov, *Intuitionistic Fuzzy Sets: Theory and Applications.* Heidelberg, Germany: Physica-Verlag, 1999.
[38] C. Cornelis, M. D. 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.
[39] C. Cornelis, G. Deschrijver, and E. E. Kerre, "Implication in intuitionistic fuzzy and interval-valued fuzzy set theory: Construction, classification, application," *Int. J. Approx. Reasoning*, vol. 35, no. 1, pp. 55–95, 2004.
[40] S. K. Samanta and T. K. Mondal, "Intuitionistic fuzzy rough sets and rough intuitionistic fuzzy sets," *J. Fuzzy Math.*, vol. 9, no. 3, pp. 561–582, 2001.
[41] L. Zhou and W. Wu, "On generalized intuitionistic fuzzy rough approximation operators," *Inf. Sci.*, vol. 178, no. 11, pp. 2448–2465, 2008.
[42] L. Zhou, W. Wu, and W. Zhang, "On characterization of intuitionistic fuzzy rough sets based on intuitionistic fuzzy implicators," *Inf. Sci.*, vol. 179, no. 7, pp. 883–898, 2009.
[43] B. Huang, Y. Zhuang, and H. Li, "Information granulation and uncertainty measures in interval-valued intuitionistic fuzzy information systems," *Eur. J. Oper. Res.*, vol. 231, no. 1, pp. 162–170, 2013.
[44] B. Huang, C. Guo, Y. Zhuang, H. Li, and X. Zhou, "Intuitionistic fuzzy multigranulation rough sets," *Inf. Sci.*, vol. 277, no. 1, pp. 299–320, 2014.
[45] B. Huang, C. Guo, H. Li, G. Feng, and X. Zhou, "Hierarchical structures and uncertainty measures for intuitionistic fuzzy approximation space," *Inf. Sci.*, vol. 336, no. 1, pp. 92–114, 2016.
[46] X. Zhang, B. Zhou, and P. Li, "A general frame for intuitionistic fuzzy rough sets," *Inf. Sci.*, vol. 216, no. 24, pp. 34–49, 2012.
[47] E. Hesameddini, M. Jafarzadeh, and H. Latifizadeh, "Rough set theory for the intuitionistic fuzzy information systems," *Int. J. Modern Math. Sci.*, vol. 6, no. 3, pp. 132–143, 2013.
[48] Z. Zhang, "Attributes reduction based on intuitionistic fuzzy rough sets," *J. Intell. Fuzzy Syst.*, vol. 30, no. 2, pp. 1127–1137, 2016.
[49] Z. Gong and X. Zhang, "Variable precision intuitionistic fuzzy rough sets model and its application," *Int. J. Mach. Learn. Cybern.*, vol. 5, no. 2, pp. 263–280, 2015.
[50] P. Jena and S. K. Ghosh, "Intuitionistic fuzzy rough sets," *Notes Intuit. Fuzzy Sets*, vol. 8, no. 1, pp. 1–18, 2002.
[51] D. Liang and D. Liu, "Deriving three-way decisions from intuitionistic fuzzy decision-theoretic rough sets," *Inf. Sci.*, vol. 300, no. 10, pp. 28–48, 2015.
[52] Y. Liu and Y. Lin, "Intuitionistic fuzzy rough set model based on conflict distance and applications," *Appl. Soft Comput.*, vol. 31, no. 2, pp. 266–273, 2015.
[53] Y. Liu, Y. Lin, and H. Zhao, "Variable precision intuitionistic fuzzy rough set model and applications based on conflict distance," *Exp. Syst.*, vol. 32, no. 2, pp. 220–227, 2015.
[54] B. Sun, W. Ma, and Q. Liu, "An approach to decision making based on intuitionistic fuzzy rough sets over two universes," *J. Oper. Res. Soc.*, vol. 64, no. 7, pp. 1079–1089, 2017.
[55] D. Coker, "Fuzzy rough sets are intuitionistic L-fuzzy sets," *Fuzzy Sets Syst.*, vol. 96, no. 3, pp. 381–383, 1998.
[56] T. Das, D. Acharjya, and M. Patr, "Multi criterion decision making using intuitionistic fuzzy rough set on two universal sets," *Int. J. Intell. Syst. Appl.*, vol. 7, no. 4, pp. 26–33, 2015.
[57] E. Szmidt and J. Kacprzyk, "Entropy for intuitionistic fuzzy sets," *Fuzzy Sets Syst.*, vol. 118, no. 3, pp. 467–477, 2001.
[58] I. K. Vlachos and G. D. Sergiadis, "Subsethood, entropy, and cardinality for interval-valued fuzzy sets-an algebraic derivation," *Fuzzy Sets Syst.*, vol. 158, no. 12, pp. 1384–1396, 2007.
[59] H. Bustince and P. Burillo, "Structures on intuitionistic fuzzy relations," *Fuzzy Sets Syst.*, vol. 78, no. 78, pp. 293–303, 1996.
[60] C. L. Blake and C. J. Merz, UCI Repository of Machine Learning Databases, 1998. [Online]. Available: http://www. ics.uci.edu/mlearn/MLRepository.html
[61] Kent Ridge Bio-medical Dataset. 2002. [Online]. Available: http://datam.i2r.a-tar.edu.sg/datasets/krbd/index.html
[62] R. Jensen and Q. Shen, "Semantics-preserving dimensionality reduction: Rough and fuzzy-rough-based approaches," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 12, pp. 1457–1471, Dec. 2004.
[63] 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.
[64] Y. Qian, Q. Wang, H. Cheng, J. Liang, and C. Dang, "Fuzzy-rough feature selection accelerator," *Fuzzy Sets Syst.*, vol. 258, no. 1, pp. 61–78, 2015.
[65] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and regression trees*, vol. 93, Belmont, CA, USA: Wadsworth, 1984, no. 99, pp. 101.


Authors' photographs and biographies not available at the time of publication.