**QUEEN'S
UNIVERSITY
BELFAST**

Fuzzy rule based classification method for incremental rule learning

Niu, J., Chen, D., Li, J., & Wang, H. (2022). Fuzzy rule based classification method for incremental rule learning. *IEEE Transactions on Fuzzy Systems*, 30(9), 3748-3761. <https://doi.org/10.1109/TFUZZ.2021.3128061>

Published in:
IEEE Transactions on Fuzzy Systems

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights
Copyright 2021, IEEE
This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Fuzzy rule based classification method for incremental rule learning

Jiaojiao Niu, Degang Chen, Jinhai Li, Hui Wang

Abstract—Granular rules have been extensively used for classification in fuzzy datasets to promote the advancement of artificial intelligence. However, due to the diversity of data types, how to improve the readability of the extracted granular rules while ensuring efficiency is always a challenge. Since granular reduct in granular computing can simplify real complex problem and dataset, this paper carries out granular rule learning from the perspective of granular reduct by taking FCA-based granular computing method as a framework. Specifically, for achieving classification task, we first propose a method to update the granular reduct, and then explore the updating mechanism of fuzzy granular rule in a reduced dataset. Secondly, a novel Fuzzy Rule based Classification Model named FRCM is presented for fuzzy granular rule learning. In order to verify the effectiveness of the proposed model, some numerical experiments for incremental learning and fuzzy rule mining are conducted to demonstrate that FRCM can achieve the state-of-the-art classification performance.

Index Terms—fuzzy concept lattice, granular reduct, fuzzy granular rule, incremental learning, rule based classification.

I. INTRODUCTION

RULE is one of the most popular types of knowledge representation approaches, which typically consists of two pieces: condition and conclusion. Normally, the former indicates the necessary condition or minimum requirement of the latter, thus rules can be used for achieving classification task. To our knowledge, many algorithms and models [1]–[8] have been proposed under the fuzzy datasets [9], [10] or relational datasets [11] for the purpose of classification learning.

Formal concept analysis (FCA) [12], as a useful tool for knowledge representation, can obtain rules through the implication relationship between concepts. And those rules can be used to achieve a variety of classification tasks. In the literatures, Kuznetsov [13] extracted classification rules (also called hypotheses) by constructing concepts from the positive and negative examples, while Nozali et al. [5] constructed a fuzzy rule-based classification system with an adaptive method for pattern classification problems. Considering that it takes

more time to get rules from concepts [12], [14]–[18], Mi et al. [19] proposed a fuzzy-based concept learning model by employing the attribute-oriented fuzzy concept similarity degree. To avoid repetitive training, researchers proposed many incremental learning systems [20]–[25] to meet different task requirements. For instance, Shi et al. [25] proposed a concurrent incremental learning technique by continuously accommodating newly added data to meet the requirements of classification task. Aiming at the non-stationary environments, Song et al. [21] introduced a window-based self-adaptive fuzzy network to continuously modify the network through identifying new knowledge from the previous samples, while Yu et al. [20] presented a Gaussian membership-based SOINN (Gm-SOINN) by employing a Gaussian membership degree to indicate which node is a winner for solving the stability-plasticity dilemma. Nakashima et al. [26] studied the incremental construction of fuzzy rule-based classifiers and Baruah et al. [27] introduced an evolving fuzzy rule-based classifier. Despite the effectiveness of those methods, they still face the challenge of large-scale datasets.

To tackle with this problem, granular computing (GrC), which is well known for its partition idea to simplify real complex problems has been successfully applied to many research fields, such as FCA [12], [28]–[33], fuzzy set [24], [34], three-way decision [35], [36] and cognitive learning [10], [32], [33], [37], [38]. Among those work, Garcia et al. [24] presented a modified evolving granular fuzzy rule-based model and Xu et al. [32] put forward a novel GrC method of machine learning from the perspective of information granules. Since the consistency [28], [29] of a formal decision context makes granular rule independent of concept lattice, the granular rule mining algorithm has linear time complexity. This makes granular rule become more efficient to be generated. Considering that granular rule is actually a kind of decision rule, and there is no work trying to examine the classification ability of the granular rule, this paper attempts to propose a classifier based on the granular rule.

Generally, granular rules are derived from granular concepts. Other concepts can be generated by certain operation between granular concepts, which makes granular rules have good granularity characteristics. Such granularity characteristics make granular rules have two advantages in classification task: (1) it is more efficient; (2) granular rule can better maintain the integrity of the acquired knowledge. At present, most of the theories of granular rule are however focused on explaining the mechanism of granular rules mining. No study to date has examined the effect of granular reduct on classification performance of granular rules. Since granular

J. Niu is with School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China (e-mail: njiao_92@163.com).

D. Chen is with School of Mathematics and Physics, North China Electric Power University, Beijing 102206, China (e-mail: wsc2973@163.com).

J. Li is with the Data Science Research Center, Kunming University of Science and Technology, Kunming 650500, China, and also with the Faculty of Science, Kunming University of Science and Technology, Kunming 650500, China (e-mail: jhlixjtu@163.com).

H. Wang is with School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, BT9 5BN Belfast, UK (e-mail: h.wang@qub.ac.uk).

reduct can lead to more compact granular rules by removing part of attributes, it is meaningful to reduce granular rules by using granular reduct technology.

Granular reduct is a kind of consistencies preserved attribute reduction methods [39]–[55], whose objective is to find a minimum attribute set to keep the pre-defined consistency unchanged. For instance, Pei et al. [54] defined the homomorphism consistent set from the view of algebra and discussed the problem of attribute reduction of concept lattice from the view of homomorphism, while Wu et al. [28] introduced the consistent set which keeps the extension of concept lattice invariant to propose granular reduct. Shao et al. [29] have proved that granular reduct and classification reduct have the same attribute characteristics in a consistent fuzzy formal decision context, which provides a strong theoretical support to employ granular reduct for purifying the datasets. Generally, granular reduct can be calculated by discernibility matrix or Boolean function [28], [29], [55]. However, due to lacking of incremental learning mechanism, they inevitably face the problem of recalculation when new objects are added, which precludes granular rule from applying effectively. Therefore, in order to reduce computation sharply, there is an urgent need for developing incremental granular reduct computing algorithms.

According to the requirements discussed above, this paper aims at simplifying granular rules by reducing the dimension of a fuzzy dataset with granular reduct and then applies them for classification task. To achieve this task, the first step is to discuss the updating mechanism of granular reduct. And then, some new notions, such as identification ability and similarity degree are defined to select the required attributes. Moreover, note that granular reduct is usually not unique and less attributes are better. We give priority to the attribute with the strongest identification ability in every update. Based on this idea, the granular rules are updated in the reduced dataset. Moreover, a Fuzzy Rule based Classification Model named FRCM is proposed. Finally, numerical experiments are conducted on 24 datasets to evaluate the classification performance of the proposed model. In summary, the main contributions of this paper are as follows:

- (1) an incremental algorithm is developed to update granular reduct;
- (2) using the granular reduct, granular rules are reduced dynamically;
- (3) we present a novel fuzzy rule based classification model.

The remainder is organized as follows. Section 2 reviews some related notions with fuzzy formal decision context, granular reduct and fuzzy granular rule. In Section 3, we first discuss the updating mechanism of granular reduct and granular rule, and then give an overall introduction to the Fuzzy Rule based Classification Model (FRCM). The numerical experiments on 24 datasets selected from UCI Machine Learning Repository and Weka follow immediately in Section 4. A summary and future work come in Section 5.

II. PRELIMINARY KNOWLEDGE

In this section, we introduce some notions and properties related to fuzzy formal decision context, granular reduct and fuzzy granular rule.

A. Fuzzy formal decision context

As is well known, a triple (U, A, I) is a formal context, where both U and A are nonempty finite sets, indicating object set and attribute set, respectively, and $I \subseteq U \times A$ is a binary relation between U and A . In general, $I(x, a) = 1$ means that the object x has the attribute a while $I(x, a) = 0$ means the opposite.

Since huge amounts of data are not 0-1 valued in real life, the binary formal context cannot effectively express them. Therefore, the fuzzy formal context was proposed based on fuzzy sets [34] to employ the membership degree $\mu_{\tilde{I}}(x, a) \in [0, 1]$ instead of $\mu_{\tilde{I}}(x, a) \in \{0, 1\}$ to show the association between the object x and the attribute a .

Let $U = \{x_1, x_2, \dots, x_n\}$ be a universe of discourse. A fuzzy set $\tilde{X}$ on U can be denoted as $\tilde{X} = \{\mu_{\tilde{X}}(x_1), \mu_{\tilde{X}}(x_2), \dots, \mu_{\tilde{X}}(x_n)\}$ or simply $\tilde{X} = \{\tilde{X}(x_1), \tilde{X}(x_2), \dots, \tilde{X}(x_n)\}$, where $\mu_{\tilde{X}}$ is a membership function mapping each $x \in U$ to a value in the unit interval $[0, 1]$. And we denote by $F(U)$ the set of all fuzzy sets on U . Then, for any $\tilde{X}, \tilde{Y} \in F(U)$, $x \in U$, $\tilde{X} \cup \tilde{Y}$, $\tilde{X} \cap \tilde{Y}$ and $\tilde{X} \subseteq \tilde{Y}$ are defined as follows:

$$\begin{aligned} (\tilde{X} \cup \tilde{Y})(x) &= \tilde{X}(x) \vee \tilde{Y}(x) = \max \{ \tilde{X}(x), \tilde{Y}(x) \}, \\ (\tilde{X} \cap \tilde{Y})(x) &= \tilde{X}(x) \wedge \tilde{Y}(x) = \min \{ \tilde{X}(x), \tilde{Y}(x) \}, \\ \tilde{X} \subseteq \tilde{Y} &\Leftrightarrow \tilde{X}(x) \leq \tilde{Y}(x) \text{ for any } x \in U. \end{aligned}$$

Definition 1: [18] A fuzzy formal context is a triple $(U, A, \tilde{I})$, where U and A are sets of objects and attributes, respectively, and $\tilde{I}$ is a fuzzy relation on $U \times A$. Each $\mu_{\tilde{I}}(x, a)$ indicates the degree of attribute a belonging to object x , which means $\tilde{I}(x, a) = \mu_{\tilde{I}}(x, a)$. And the larger the value $\mu_{\tilde{I}}(x, a)$ is, the bigger the belonging degree is.

The classical formal context can be regarded as a special fuzzy formal context, that is, the value of the membership degree is 0 or 1. In what follows, **we denote by $P(U)$ the set of all ordinary subsets of U** . Then, we can define the following fuzzy operators.

Definition 2: [15], [16], [56] Let $(U, A, \tilde{I})$ be a fuzzy formal context. For $X \subseteq P(U)$, $\tilde{B} \in F(A)$, the operators $f : P(U) \rightarrow F(A)$ and $g : F(A) \rightarrow P(U)$ are defined respectively as follows:

$$\begin{aligned} f(X)(a) &= \bigwedge_{x \in X} \tilde{I}(x, a), a \in A, \\ g(\tilde{B}) &= \{x \in U | \forall a \in A, \tilde{B}(a) \leq \tilde{I}(x, a)\}. \end{aligned} \tag{1}$$

Property 1: [15] Let $(U, A, \tilde{I})$ be a fuzzy formal context, $X, Y \in P(U)$ and $\tilde{B}, \tilde{C} \in F(A)$. Then

- (1) $X \subseteq Y \Rightarrow f(Y) \subseteq f(X), \tilde{B} \subseteq \tilde{C} \Rightarrow g(\tilde{C}) \subseteq g(\tilde{B})$;
- (2) $X \subseteq g \circ f(X), \tilde{B} \subseteq f \circ g(\tilde{B})$;
- (3) $f(X) = f \circ g \circ f(X), g(\tilde{B}) = g \circ f \circ g(\tilde{B})$.

From Property 1, we can conclude that: (1) the membership degree of the whole Y is less than that of the part X ; (2) the compound operation maps an object set to a larger one, and makes the membership function of a fuzzy set bigger; (3)

$f(X)$ and $g(\tilde{B})$ can describe the object set X and the fuzzy set $\tilde{B}$ completely.

An object set X (extent) and a description fuzzy set $\tilde{B}$ (intent) satisfying $f(X) = \tilde{B}$ and $g(\tilde{B}) = X$ can form a pair $(X, \tilde{B})$, called a fuzzy concept. Since X is crisp while $\tilde{B}$ is fuzzy, such a fuzzy concept was also called crisp-fuzzy concept in [15] and [16]. In particular, $(g \circ f(x), f(x))$ for each $x \in U$ are fuzzy object concepts, called granular concepts. For two fuzzy concepts $(X_1, \tilde{B}_1)$ and $(X_2, \tilde{B}_2)$, we define the following partial order:

$$(X_1, \tilde{B}_1) \leq (X_2, \tilde{B}_2) \Leftrightarrow X_1 \subseteq X_2 (\tilde{B}_2 \subseteq \tilde{B}_1). \quad (2)$$

For convenience, we use $L(U, A, \tilde{I})$ to denote the set of all fuzzy concepts of $(U, A, \tilde{I})$. Then the infimum and the supremum can be defined as:

$$\begin{aligned} (X_1, \tilde{B}_1) \wedge (X_2, \tilde{B}_2) &= (X_1 \cap X_2, f \circ g(\tilde{B}_1 \cup \tilde{B}_2)) \\ &= (X_1 \cap X_2, f(X_1 \cap X_2)), \\ (X_1, \tilde{B}_1) \vee (X_2, \tilde{B}_2) &= (g \circ f(X_1 \cup X_2), \tilde{B}_1 \cap \tilde{B}_2) \\ &= (g(\tilde{B}_1 \cap \tilde{B}_2), \tilde{B}_1 \cap \tilde{B}_2). \end{aligned} \quad (3)$$

Based on two fuzzy formal contexts, a fuzzy formal decision context can be defined to meet the needs of decision making.

Definition 3: [57] A fuzzy formal decision context is a quintuple $S = (U, A, \tilde{I}, D, \tilde{J})$, where $(U, A, \tilde{I})$ and $(U, D, \tilde{J})$ are fuzzy formal contexts with $A \cap D = \emptyset$, and A, D are conditional attribute set and decision attribute set, respectively.

A quintuple $S = (U, A, \tilde{I}, D, \tilde{J})$ is known as a crisp-fuzzy formal decision context in [57], while $S = (U, A, \tilde{I}, D, J)$ was called a fuzzy-crisp formal decision context in [19]. The latter satisfying regularity is widely used in classification tasks, and a regular fuzzy-crisp formal decision context is a quintuple $S = (U, A, \tilde{I}, D, J)$, in which $g(a) \cap g(b) = \emptyset$ holds for any $a, b \in D$. Since a classical formal context can also be regarded as a kind of special fuzzy formal context, both crisp-fuzzy formal decision context and fuzzy-crisp formal decision context are fuzzy formal decision contexts. Besides, in order to distinguish the operators f and g in two fuzzy formal contexts, we give different subscripts to them in different fuzzy formal contexts. That is, for $C \subseteq A$, f_C and g_C in the fuzzy formal context $S = (U, C, \tilde{I}_C)$ are defined as:

$$\begin{aligned} f_C(X)(a) &= \bigwedge_{x \in X} \tilde{I}(x, a), a \in C, \\ g_C(\tilde{B}) &= \{x \in U | \forall a \in C, \tilde{B}(a) \leq \tilde{I}_C(x, a)\}, \tilde{B} \subseteq F(C). \end{aligned} \quad (4)$$

To avoid confusion, $f_{EC}(X)$ and $g_{EC}(\tilde{B})$ for any $E \subseteq U$ are employed to denote the operates f and g in $(E, C, \tilde{I}_{EC})$ when completely necessary.

Example 1: Table I is a fuzzy formal decision context $S = (U, A, \tilde{I}, D, J)$, where $(U, A, \tilde{I})$ is a fuzzy formal context and (U, D, J) is a formal context.

TABLE I: A Fuzzy Formal Decision Context
 $S = (U, A, \tilde{I}, D, J)$

U	a_1	a_2	a_3	a_4	a_5	a_6	d_1	d_2
x_1	0	0.19	0.73	0.03	0.99	0	0	1
x_2	0.84	0.26	0	0	0.65	0.93	0	1
x_3	0.97	0.40	0	0	0.89	0.40	1	0
x_4	0.78	0.23	0	0	0.87	0.49	1	0
x_5	0	0	0	0	0.86	0.52	1	0

The fuzzy formal context $(U, A, \tilde{I})$ has 17 concepts:

$$\begin{aligned} &(\{x_1, x_2, x_3, x_4, x_5\}, \{a_1^0, a_2^0, a_3^0, a_4^0, a_5^{0.65}, a_6^0\}), \\ &(\{x_2, x_3, x_4, x_5\}, \{a_1^0, a_2^0, a_3^0, a_4^0, a_5^{0.65}, a_6^{0.40}\}), \\ &(\{x_1, x_2, x_3, x_4\}, \{a_1^0, a_2^{0.19}, a_3^0, a_4^0, a_5^{0.65}, a_6^0\}), \\ &(\{x_1, x_3, x_4, x_5\}, \{a_1^0, a_2^0, a_3^0, a_4^0, a_5^{0.86}, a_6^0\}), \\ &(\{x_3, x_4, x_5\}, \{a_1^0, a_2^0, a_3^0, a_4^0, a_5^{0.86}, a_6^{0.40}\}), \\ &(\{x_1, x_3, x_4\}, \{a_1^0, a_2^{0.19}, a_3^0, a_4^0, a_5^{0.87}, a_6^0\}), \\ &(\{x_4, x_5\}, \{a_1^0, a_2^0, a_3^0, a_4^0, a_5^{0.86}, a_6^{0.49}\}), \\ &(\{x_3, x_4\}, \{a_1^{0.78}, a_2^{0.23}, a_3^0, a_4^0, a_5^{0.87}, a_6^{0.40}\}), \\ &(\{x_2, x_4\}, \{a_1^{0.78}, a_2^{0.23}, a_3^0, a_4^0, a_5^{0.65}, a_6^{0.49}\}), \\ &(\{x_2, x_3\}, \{a_1^{0.84}, a_2^{0.26}, a_3^0, a_4^0, a_5^{0.65}, a_6^{0.40}\}), \\ &(\{x_1, x_3\}, \{a_1^0, a_2^{0.19}, a_3^0, a_4^0, a_5^{0.89}, a_6^0\}), \\ &(\{x_1\}, \{a_1^0, a_2^{0.19}, a_3^{0.73}, a_4^{0.03}, a_5^{0.99}, a_6^0\}), \\ &(\{x_2\}, \{a_1^{0.84}, a_2^{0.26}, a_3^0, a_4^0, a_5^{0.65}, a_6^{0.93}\}), \\ &(\{x_3\}, \{a_1^{0.97}, a_2^{0.40}, a_3^0, a_4^0, a_5^{0.89}, a_6^{0.40}\}), \\ &(\{x_4\}, \{a_1^{0.78}, a_2^{0.23}, a_3^0, a_4^0, a_5^{0.87}, a_6^{0.49}\}), \\ &(\{x_5\}, \{a_1^0, a_2^0, a_3^0, a_4^0, a_5^{0.86}, a_6^{0.52}\}), \\ &(\emptyset, \{a_1^{0.97}, a_2^{0.40}, a_3^{0.73}, a_4^{0.03}, a_5^{0.99}, a_6^{0.93}\}). \end{aligned}$$

The formal context (U, D, J) has 4 concepts:

$$\begin{aligned} &(\{x_1, x_2\}, \{d_2\}), (\{x_3, x_4, x_5\}, \{d_1\}), \\ &(\{x_1, x_2, x_3, x_4, x_5\}, \emptyset), (\emptyset, \{d_1, d_2\}). \end{aligned}$$

B. Granular reduct and fuzzy granular rule

Granular reduct refers to a minimum attribute set to maintain the consistency of a formal decision context. We first recall some related notions.

Definition 4: [29] Let $S = (U, A, \tilde{I}, D, J)$ be a fuzzy formal decision context. S is said to be consistent if $g_A \circ f_A(x) \subseteq g_D \circ f_D(x)$ for all $x \in U$. Otherwise, it is said to be inconsistent.

Example 2: The fuzzy formal decision context $S = (U, A, \tilde{I}, D, J)$ is consistent because:

$$\begin{aligned} g_A \circ f_A(x_1) &= \{x_1\} \subseteq g_D \circ f_D(x_1) = \{x_1, x_2\}, \\ g_A \circ f_A(x_2) &= \{x_2\} \subseteq g_D \circ f_D(x_2) = \{x_1, x_2\}, \\ g_A \circ f_A(x_3) &= \{x_3\} \subseteq g_D \circ f_D(x_3) = \{x_3, x_4, x_5\}, \\ g_A \circ f_A(x_4) &= \{x_4\} \subseteq g_D \circ f_D(x_4) = \{x_3, x_4, x_5\}, \\ g_A \circ f_A(x_5) &= \{x_5\} \subseteq g_D \circ f_D(x_5) = \{x_3, x_4, x_5\}. \end{aligned}$$

Definition 5: [29] Let $S = (U, A, \tilde{I}, D, J)$ be a consistent fuzzy formal decision context and $C \subseteq A$. If $g_C \circ f_C(x) \subseteq g_D \circ f_D(x)$ for all $x \in U$, then C is referred to as a granular consistent set of S . If C is a granular consistent set of S and no proper subset of C is a granular consistent set, then C is called a granular reduct of S .

Obviously, for a consistent fuzzy formal decision context, its granular reduct may no longer be a consistent set if some attributes are reduced. On the other hand, since the granular reduct is the minimum attribute set to maintain consistency,

adding extra attributes will not affect its consistency. Therefore, the following property holds:

Property 2: Let $S = (U, A, \tilde{I}, D, J)$ be a consistent fuzzy formal decision context, $C \subseteq A$ is a granular reduct of S and $E = A - C$. Then, for $E^* \subseteq E$, $g_{C \cup E^*} \circ f_{C \cup E^*}(x) \subseteq g_D \circ f_D(x)$.

Proof It follows immediately from Property 1 and Definition 5.

Similar to the classical concept lattice theory, a rule can also be extracted from the granular concepts $(g_A \circ f_A(x), f_A(x))$ and $(g_D \circ f_D(x), f_D(x))$ in $(U, A, \tilde{I}, D, J)$ if $g_A \circ f_A(x) \subseteq g_D \circ f_D(x)$. Since the consistency of the fuzzy formal decision context ensures the condition of rule generation, we can obtain a rule for each $x \in U$.

Definition 7: Let $S = (U, A, \tilde{I}, D, J)$ be a consistent fuzzy formal decision context. $(g_A \circ f_A(x), f_A(x))$ and $(g_D \circ f_D(x), f_D(x))$ are two fuzzy granular concepts in $(U, A, \tilde{I})$ and (U, D, J) , respectively. Then, $r_x : f_A(x) \rightarrow f_D(x)$ is called a fuzzy granular rule of S , where $f_A(x)$ and $f_D(x)$ are called condition and conclusion of r_x , respectively.

For a consistent fuzzy formal decision context $S = (U, A, \tilde{I}, D, J)$, we denote $R(S)$ as the set of all the fuzzy granular rules. Especially, the conclusion $f_D(x)$ in r_x is also marked as $label(r_x)$ or $label(x)$ for convenience.

III. FUZZY RULE UPDATING METHOD

In this section, we aim at presenting a novel Fuzzy Rule based Classification Model (FRCM). To achieve this goal, we first introduce how to obtain a new granular reduct from the original one when a new object is added into a fuzzy formal decision context, and then explore the updating mechanism of granular rules. In order to distinguish the updated fuzzy formal decision context from the previous one, $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ is employed to denote S after adding a new object x^* , i.e. the updated version of S . Similarly, the operators f and g in Definition 2 for S^* are remarked by f^* and g^* , respectively.

A. The granular reduct updating mechanism

In this subsection, we introduce an updating mechanism for computing granular reduct. According to Definition 5, if the granular reduct C of S is no longer the granular consistent set of S^* , it means that there exists $x \in U^*$ such that it does not satisfy the consistency. In other words, the newly added object x^* makes the object set U be divided into two disjoint subsets U_1 and U_2 , in which for any $x_1 \in U_1$, $g_C^* \circ f_C^*(x_1) \subseteq g_D^* \circ f_D^*(x_1)$ and for any $x_2 \in U_2$, $g_C^* \circ f_C^*(x_2) \not\subseteq g_D^* \circ f_D^*(x_2)$. To update the granular reduct, we need to know what impact does x^* brings to $g_C \circ f_C(x)$.

Theorem 1: Let $S = (U, A, \tilde{I}, D, J)$ and $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ be two consistent fuzzy formal decision contexts, and $C \subseteq A$. Then, for any $x \in U$, we have

$$g_C^* \circ f_C^*(x) = \begin{cases} g_C \circ f_C(x) \cup \{x^*\}, & \text{if } f_C(x) \subseteq f_C^*(x^*), \\ g_C \circ f_C(x), & \text{otherwise.} \end{cases}$$

TABLE II: A comparison of $f(x)$ with $f^*(x^*)$

Case	$(U, A, \tilde{I})$	(U, D, J)
(1)	$f(x) \subset f^*(x^*)$	$f_D(x) = f_D^*(x^*)$
(2)	$f(x) \supset f^*(x^*)$	$f_D(x) \neq f_D^*(x^*)$
(3)	$f(x) = f^*(x^*)$	
(4)	else	

Proof: It is obvious from the definitions of S and S^* , as well as Definition 2.

Theorem 1 indicates that a new added object may increase the extent of the original granular concept. Since reduce part of attributes will makes extent larger, the consistency of a fuzzy formal decision context involves additional attributes. According to Property 2, adding additional attributes to the granular reduct does not affect its consistency. To obtain a minimum granular consistent set, we need to find a minimum attribute set $C_+ \subseteq A - C$ which satisfies $g_{C \cup C_+}^* \circ f_{C \cup C_+}^*(x) \subseteq g_D^* \circ f_D^*(x)$ for any x that is inconsistent with consistency.

To obtain such an attribute set C_+ , two questions must be answered: one is how to select attributes needed while the other is how to ensure the minimum. For our purpose, a comparison of $f(x), x \in U$ and $f^*(x^*), x^* \in U^*$ is shown in Table II. As Table II illustrates, it has four cases in terms of relationship between $f(x)$ and $f^*(x^*)$ in $(U, A, \tilde{I})$, in which “else” indicates that it does not belong to the first three cases. And since the decision attribute only contains label information in classification task, there are only two cases in (U, D, J) . It is the inconsistent change of the relationship between $g_C^* \circ f_C^*(x)$ and $g_D^* \circ f_D^*(x)$ that leads to the fact that C is no longer a granular consistent set of S^* . In what follows, we discuss how to update granular reduct by adding additional attributes according to different situations listed in Table II.

Theorem 2: Let $S = (U, A, \tilde{I}, D, J)$ and $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ be two consistent fuzzy formal decision contexts, and C be a granular reduct of S . C is a granular reduct of S^* if any $x \in U$ satisfies one of the following conditions:

- (1) $f_D(x) = f_D^*(x)$,
- (2) $f_D(x) \neq f_D^*(x^*), \text{ else,}$

where “else” indicates that $f_C(x)$ and $f_C^*(x^*)$ satisfy Case (4) in Table II.

Proof: Suppose any $x \in U$ satisfies (1) and (2), then

$$\begin{aligned} g_C^* \circ f_C^*(x) &\subseteq g_C \circ f_C(x) \cup \{x^*\}, \\ g_D^* \circ f_D^*(x) &\subseteq g_D \circ f_D(x) \cup \{x^*\}. \end{aligned}$$

Since $g_C \circ f_C(x) \subseteq g_D \circ f_D(x)$, then $g_C^* \circ f_C^*(x) \subseteq g_D^* \circ f_D^*(x)$. Moreover, $g_C^* \circ f_C^*(x^*) \subseteq g_C \circ f_C(x) \cup \{x^*\} \subseteq g_D \circ f_D(x) \cup \{x^*\} = g_D^* \circ f_D^*(x^*)$ for item (1) and $g_C^* \circ f_C^*(x^*) = \{x^*\} = g_D^* \circ f_D^*(x^*)$ for item (2). Thus, C is a granular consistent of S^* .

Moreover, since C is a granular reduct of S , for any $C^* \subseteq C$, there must exist $x \in U$, $g_{C^*} \circ f_{C^*}(x) \not\subseteq g_D \circ f_D(x)$ also $g_{C^*}^* \circ f_{C^*}^*(x) \not\subseteq g_D^* \circ f_D^*(x)$. Hence, C is a granular reduct of S^* .

Theorem 2 shows that if the new object has the same label with the original objects, then C is also a granular reduct of S^* . Otherwise, C is a granular reduct of S^* if and only if $g_C \circ f_C(x)$ is invariant in S^* for any $x \in U$ and $g_C^* \circ f_C^*(x^*) \cap U = \emptyset$.

Theorem 3: Let $S = (U, A, \tilde{I}, D, J)$ and $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ be two consistent fuzzy formal decision contexts, and C be a granular reduct of S . $C^* = C \cup \{a\}$, $a \in A - C$ with $\tilde{I}(x, a) \subset \tilde{I}^*(x^*, a)$ is a granular consistent set of S^* if $f_D(x) \neq f_D^*(x^*)$, $f_C(x) \supset f_C^*(x^*)$ for any $x \in U$.

Proof: If any $x \in U$ can satisfy $f_D(x) \neq f_D^*(x^*)$, $f_C(x) \supset f_C^*(x^*)$. Due to Property 2 and Theorem 1, $x \in g_C^* \circ f_C^*(x^*)$, but $x \notin g_D^* \circ f_D^*(x^*)$. Then, $g_C^* \circ f_C^*(x^*) \not\subseteq g_D^* \circ f_D^*(x^*)$, which lead to the fact that C is not a granular consistent set of S^* .

Since $\tilde{I}(x, a) \subset \tilde{I}^*(x^*, a)$, then $f_C^*(x) \cap f_C^*(x^*)$ belongs to Case (4) in Table II. Thus, $g_{C^*}^* \circ f_{C^*}^*(x) = g_C^* \circ f_C^*(x) \subseteq g_D^* \circ f_D(x) = g_D^* \circ f_D^*(x)$. Furthermore, $g_{C^*}^* \circ f_{C^*}^*(x^*) = \{x^*\} \subseteq g_D^* \circ f_D^*(x^*) = \{x^*\}$. In summary, C^* is a granular consistent set of S^* .

Theorem 4: Let $S = (U, A, \tilde{I}, D, J)$ and $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ be two consistent fuzzy formal decision contexts, and C be a granular reduct of S . $C^* = C \cup \{a\}$ with $\tilde{I}(x, a) \supset \tilde{I}^*(x^*, a)$, $a \in A - C$ is a granular consistent set of S^* if $f_D(x) \neq f_D^*(x^*)$, $f_C(x) \subset f_C^*(x^*)$ for any $x \in U$.

Proof: Suppose any $x \in U$ can satisfy $f_D(x) \neq f_D^*(x^*)$, $f_C(x) \subset f_C^*(x^*)$. According to Property 2 and Theorem 1, $g_C^* \circ f_C^*(x) = g_C \circ f_C(x) \cup \{x^*\}$ and $g_D^* \circ f_D^*(x) = g_D \circ f_D(x)$. Thus, $g_C^* \circ f_C^*(x^*) \not\subseteq g_D^* \circ f_D^*(x^*)$, C is not a granular consistent set of S^* .

For any $a \in A - C$, if $\tilde{I}(x, a) \supset \tilde{I}^*(x^*, a)$, then $f_C^*(x) \cap f_C^*(x^*)$ belongs to Case (4) in Table II. According to Theorem 1, we obtain $g_{C^*}^* \circ f_{C^*}^*(x) = g_C^* \circ f_C^*(x) \subseteq g_D^* \circ f_D(x) = g_D^* \circ f_D^*(x)$. Furthermore, $g_{C^*}^* \circ f_{C^*}^*(x^*) = \{x^*\} \subseteq g_D^* \circ f_D^*(x^*) = \{x^*\}$. So, C^* is a granular consistent set of S^* .

Theorem 5: Let $S = (U, A, \tilde{I}, D, J)$ and $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ be two consistent fuzzy formal decision contexts, and C be a granular reduct of S . $C^* = C \cup \{a, b\}$ with $\tilde{I}(x, a) \supset \tilde{I}^*(x^*, a)$, $\tilde{I}(x, b) \subset \tilde{I}^*(x^*, b)$, $a, b \in A - C$ is a granular consistent set of S^* if $f_D(x) \neq f_D^*(x^*)$, $f_C(x) = f_C^*(x^*)$ for any $x \in U$.

Proof: Assume that any $x \in U$ can satisfy $f_D(x) \neq f_D^*(x^*)$, $f_C(x) = f_C^*(x^*)$. Because of Property 2 and Theorem 1, $g_C^* \circ f_C^*(x) = g_C \circ f_C(x) \cup \{x^*\}$ and $g_D^* \circ f_D^*(x) = g_D \circ f_D(x)$. Thus, $g_C^* \circ f_C^*(x^*) \not\subseteq g_D^* \circ f_D^*(x^*)$, which reveals that C is not a granular consistent set of S^* .

For any $a, b \in A - C$, if $\tilde{I}(x, a) \supset \tilde{I}^*(x^*, a)$, $\tilde{I}(x, b) \subset \tilde{I}^*(x^*, b)$, then $f_C^*(x) \cap f_C^*(x^*)$ belongs to Case (4) in Table II. According to Property 1 and Theorem 1, we get $g_{C^*}^* \circ f_{C^*}^*(x) = g_C^* \circ f_C^*(x) \subseteq g_D^* \circ f_D(x) = g_D^* \circ f_D^*(x)$. In addition, $g_{C^*}^* \circ f_{C^*}^*(x^*) = \{x^*\} \subseteq g_D^* \circ f_D^*(x^*) = \{x^*\}$. In summary, C^* is a granular consistent set of S^* .

According to the discussion of Theorems 3-5, we know that one or two additional attributes can prevent $g_C \circ f_C(x)$ from becoming larger, so that C is also a consistent set in S^* . However, due to the complexity of data, the situations discussed in Theorem 2-5 may exist at the same time after a new object is added. Additional attributes are needed for each object that do not meet the consistency. Therefore, we

need to prove that all the added attributes together with C is a consistent set of S^* .

Theorem 6: Let $S = (U, A, \tilde{I}, D, J)$ and $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ be two consistent fuzzy formal decision contexts, and C be a granular reduct of S . Let $C_+ \subseteq A - C$ be the newly added attributes according to Theorems 3-5, then $C^* = C \cup C_+$ is a granular consistent set of S^* .

Proof: We first divide U into U_1, U_2, U_3, U_4 and denote by $E_2, E_3, E_4 \subseteq C_+$ the added attributes by U_2, U_3, U_4 . We only need to prove that for $x \in U$, $g_{C^*}^* \circ f_{C^*}^*(x) \subseteq g_D^* \circ f_D^*(x)$ and for $x^* \in U^*$, $g_{C^*}^* \circ f_{C^*}^*(x^*) \subseteq g_D^* \circ f_D^*(x^*)$.

It is obvious that for $x \in U_1$, $g_C^* \circ f_C^*(x) \subseteq g_D^* \circ f_D^*(x)$ and for $x \in U_i (i = 2, 3, 4)$, $g_{C \cup E_i}^* \circ f_{C \cup E_i}^*(x) \subseteq g_D^* \circ f_D^*(x)$. According to Property 1 and Theorem 2, for $x \in U_1$ and $x \in U_i (i = 2, 3, 4)$, we obtain $g_{C^*}^* \circ f_{C^*}^*(x) \subseteq g_D^* \circ f_D^*(x)$ and $g_{U_i(C \cup E_2 \cup E_3 \cup E_4)}^* \circ f_{U_i(C \cup E_2 \cup E_3 \cup E_4)}^*(x) \subseteq g_{U_i D}^* \circ f_{U_i D}^*(x)$. Thus, for $x \in U$, $g_{C^*}^* \circ f_{C^*}^*(x) = g_{C \cup E_2 \cup E_3 \cup E_4}^* \circ f_{C \cup E_2 \cup E_3 \cup E_4}^*(x) \subseteq g_D^* \circ f_D^*(x)$.

Besides, for $x^* \in U^*$, $g_{C^*}^* \circ f_{C^*}^*(x^*) = g_{U_1 C^*} \circ f_{U_1 C^*}(x^*) \cup g_{U_2 C^*} \circ f_{U_2 C^*}(x^*) \cup g_{U_3 C^*} \circ f_{U_3 C^*}(x^*) \cup g_{U_4 C^*} \circ f_{U_4 C^*}(x^*) \cup \{x^*\} \subseteq g_{U_1 D} \circ f_{U_1 D}(x^*) \cup g_{U_2 D} \circ f_{U_2 D}(x^*) \cup g_{U_3 D} \circ f_{U_3 D}(x^*) \cup g_{U_4 D} \circ f_{U_4 D}(x^*) \cup \{x^*\} = g_D^* \circ f_D^*(x^*)$.

Up to now, we have answered how to select the required attributes. In order to ensure the minimum, metrics to measure the importance of attributes should be set up to filter out redundant attributes during the update process. Therefore, some additional notions need to be introduced.

Definition 9: Let $S = (U, A, \tilde{I}, D, J)$ be a consistent fuzzy formal decision context. For $x \in U$ and $a \in A$, we define the positive identification set (PIS) and negative identification set (NIS) of a to x on $\dot{U} \subseteq U$ as

$$\begin{aligned} \text{PIS}_{\dot{U}}(x, a) &= \{y \in \dot{U} | \tilde{I}(y, a) > \tilde{I}(x, a)\}, \\ \text{NIS}_{\dot{U}}(x, a) &= \{y \in \dot{U} | \tilde{I}(y, a) < \tilde{I}(x, a)\}. \end{aligned} \quad (5)$$

Furthermore, we define the corresponding positive identification degree (PID) and negative identification degree (NID) as

$$\begin{aligned} \text{PID}_{\dot{U}}(x, a) &= \frac{|\text{PIS}_{\dot{U}}(x, a)|}{|\dot{U}|}, \\ \text{NID}_{\dot{U}}(x, a) &= \frac{|\text{NIS}_{\dot{U}}(x, a)|}{|\dot{U}|}. \end{aligned} \quad (6)$$

Theorem 7: Let $S = (U, A, \tilde{I}, D, J)$ and $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ be two consistent fuzzy formal decision contexts, C be a granular reduct of S and $C_+ \subseteq A - C$ be the newly added attributes. For $a \in C_+$, $C \cup C_+ - \{a\}$ is a granular consistent set of S^* if and only if $\text{PIS}_{U-U_1}(x^*, a) \cup \text{NIS}_{U-U_1}(x^*, a) \subseteq \bigcup_{e \in C_+ - \{a\}} (\text{PIS}_{U-U_1}(x^*, e) \cup \text{NIS}_{U-U_1}(x^*, e))$, where U_1 is the object set satisfying the conditions in Theorem 2.

Proof: It is obvious from Theorem 6.

With the definition of the identification degree, the attribute with the strongest identification ability will be selected in each step of the updating process to minimize the number of required attributes. This idea is shown with Steps 2-8 in Algorithm 1. Considering that part of attributes may also be redundant to granular reduct, we still need to eliminate

such redundant attributes. The minimum of the new added attribute set makes redundant attributes concentrated in the original granular reduct. Therefore, we can draw the following conclusion.

Algorithm 1 The updating-based granular reduct computing method

Input: Object sets U_1, U_2, U_3, U_4 satisfying Theorems 2-5, a granular reduct C and a new object x^* .

Output: The updated granular reduct C^* .

```

1: Initialize  $\overleftarrow{U} = U_2 \cup U_4$ ,  $\overrightarrow{U} = U_3 \cup U_4$ ,  $U_* = \bigcup_{i=2}^4 U_i$ ,
    $C_+ = \emptyset$  and  $C_- = \emptyset$ ;
2: While ( $U_* \neq \emptyset$ )
3:   Compute  $\text{PIS}_{\overleftarrow{U}}(x^*, a)$  and  $\text{NIS}_{\overleftarrow{U}}(x^*, a)$  for each  $a$  in
    $A - C$ ;
4:   Select attribute  $\dot{a}$  with  $\text{PID}_{\overleftarrow{U}}(x^*, \dot{a}) + \text{NID}_{\overleftarrow{U}}(x^*, \dot{a}) =$ 
    $\max(\text{PID}_{\overleftarrow{U}}(x^*, a) + \text{NID}_{\overleftarrow{U}}(x^*, a))$ ;
5:    $C_+ = C_+ \cup \{\dot{a}\}$ ;
6:    $\overrightarrow{U} = \overrightarrow{U} - \{x \in \overrightarrow{U} | \tilde{I}(x, \dot{a}) > \tilde{I}^*(x^*, \dot{a})\}$ ;
7:    $\overleftarrow{U} = \overleftarrow{U} - \{x \in \overleftarrow{U} | \tilde{I}(x, \dot{a}) < \tilde{I}^*(x^*, \dot{a})\}$ ;
8: End
9: For each  $b$  in  $C_+$ 
10:  If  $\text{PIS}_{U_*}(x^*, b) \cup \text{NIS}_{U_*}(x^*, b) \subseteq \bigcup_{e \in E_*} (\text{PIS}_{U_*}(x^*, e) \cup$ 
    $\text{NIS}_{U_*}(x^*, e))$ , in which  $E_* = C_+ - \{b\}$ 
11:     $C_+ = C_+ - \{b\}$ ;
12:  End
13: End
14: For each  $a$  in  $C$ 
15:  If  $C \cup C_+ - \{a\}$  is a granular consistent set
16:     $C_- = C_- \cup \{a\}$ ;
17:  End
18: End
19:  $C^* = C \cup C_+ - C_-$ ;
20: Return  $C^*$ .
```

Theorem 8: Let $S = (U, A, \tilde{I}, D, J)$ and $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ be two consistent fuzzy formal decision contexts, C be a granular reduct of S . If $C^* = C \cup C_+ - \{a\}$ be a granular consistent set of S^* , then $a \in C$.

Proof: It is obvious from discussion above.

On the basis of the above discussion, we know that the updating of granular reduct involves two parts: one is to find a minimum set of attribute set **from** $A - C$, which can identify all objects that do not conform to the consistency; the other is to remove a largest attribute set from C that are redundant for granular reduct.

B. The granular rule updating mechanism

With the updating of granular reduct, the extracted fuzzy granular rules also need to be adjusted to meet the needs of the new fuzzy formal decision context. Therefore, this section mainly discuss the mechanism of updating fuzzy granular rules. Since the conclusion of a fuzzy granular rule is invariable, the key of this problem lies in updating the condition.

Theorem 9: Let $S = (U, A, \tilde{I}, D, J)$ and $S^* = (U^*, A, \tilde{I}^*, D, J^*)$ be two consistent fuzzy formal decision contexts, and $C, C^* = C \cup C_+$ be their granular reducts, respectively. Then, for a fuzzy granular rule $r_x : f_{C^*}^*(x) \rightarrow f_D^*(x)$ in $R(S^*)$, we have

$$f_{C^*}^*(x) = \begin{cases} f_C(x), & \text{if } C_+ = \emptyset, \\ f_C(x) \cup f_{C_+}(x), & \text{otherwise.} \end{cases}$$

Proof: It is obvious from Definition 7 and Theorem 8.

Theorem 9 indicates that the updating of fuzzy granular rule also involves two parts: one is to update the original rules, the other is to add the new fuzzy granular rule. We show the whole updating process in Algorithm 2 and present an example to illustrate the updating of granular reduct as well as granular rules.

Algorithm 2 The updating algorithm of fuzzy granular rules

Input: Initial fuzzy formal decision context $S = (U, A, \tilde{I}, D, J)$, a granular reduct C of S and a new object x^* .

Output: The fuzzy granular rule set $R(S^*)$.

```

1: Initialize fuzzy granular rule set  $R(S)$  of  $S$ ;
2: Initialize  $R(S^*) = \emptyset$ ;
3: Get  $C^*$  from Algorithm 1;
4: If  $C^* = C$ 
5:    $R(S^*) \leftarrow R(S) \cup \{r_{x^*}\}$ ;
6: Else
7:   For each  $r_x$  in  $R(S)$ 
8:      $f_C(x) \leftarrow f_C(x) \cup f_{C^* - C}(x)$ ;
9:      $R_{S^*} \leftarrow \{r_x\}$ ;
10:  End
11:   $R(S^*) \leftarrow \{r_{x^*}\}$ ;
12: End
13: Return  $R(S^*)$ .
```

Example 3: We take Table I as an example to show the updating process of granular reduct and granular rule. And suppose that $U_0 = \{x_1, x_2\}$ is the initial object and $U^+ = \{x_3, x_4, x_5\}$ is the incremental object set.

We first initiate the granular reduct $C = \{a_1, a_3\}$ and granular rules as:

$$\begin{aligned} r_{x_1} : \left\{ \frac{0}{a_1}, \frac{0.73}{a_3} \right\} &\rightarrow \{d_2\}, \\ r_{x_2} : \left\{ \frac{0.84}{a_1}, \frac{0}{a_3} \right\} &\rightarrow \{d_2\}. \end{aligned}$$

(1) For the new object x_3 , $U = \{x_1, x_2\}$, $U^* = \{x_1, x_2, x_3\}$. Since $f_C(x_1)$ and $f_C^*(x_3)$ satisfy Theorem 2 and $f_C(x_2) \subset f_C^*(x_3)$ satisfies Theorem 4, which leads to the fact that $\overleftarrow{U} = \{x_2\}$ and $\overrightarrow{U} = \emptyset$. According to Equation (5),

$$\begin{aligned} \text{PIS}_{\overleftarrow{U}}(x_3, a_2) &= \emptyset; & \text{NIS}_{\overleftarrow{U}}(x_3, a_2) &= \emptyset; \\ \text{PIS}_{\overleftarrow{U}}(x_3, a_4) &= \emptyset; & \text{NIS}_{\overleftarrow{U}}(x_3, a_2) &= \emptyset; \\ \text{PIS}_{\overleftarrow{U}}(x_3, a_5) &= \emptyset; & \text{NIS}_{\overleftarrow{U}}(x_3, a_2) &= \emptyset; \\ \text{PIS}_{\overleftarrow{U}}(x_3, a_6) &= \{x_2\}; & \text{NIS}_{\overleftarrow{U}}(x_3, a_2) &= \emptyset. \end{aligned}$$

According to Equation (6), $\text{PID}_{\overleftarrow{U}}(x_3, a) = \frac{|\text{PIS}_{\overleftarrow{U}}(x_3, a)|}{|\overleftarrow{U}|}$ and

$NID_{\vec{U}}(x_3, a) = \frac{|NIS_{\vec{U}}(x_3, a)|}{|\vec{U}|}$. Thus

$$\begin{aligned} PID_{\vec{U}}(x_3, a_2) &= 0; NID_{\vec{U}}(x_3, a_2) = 0; \\ PID_{\vec{U}}(x_3, a_4) &= 0; NID_{\vec{U}}(x_3, a_4) = 0; \\ PID_{\vec{U}}(x_3, a_5) &= 0; NID_{\vec{U}}(x_3, a_5) = 0; \\ PID_{\vec{U}}(x_3, a_6) &= \frac{1}{2}, NID_{\vec{U}}(x_3, a_6) = 0, \end{aligned}$$

then $C_1 = C \cup \{a_6\} = \{a_1, a_3, a_6\}$ is a granular consistent set of the updated fuzzy formal decision context. By removing each attribute in C in turns, we know that $C_1^* = C_1$ is a granular reduct. The updated granular rules are as follows:

$$\begin{aligned} r_{x_1}^1 &: \left\{ \frac{0}{a_1}, \frac{0.73}{a_3}, \frac{0}{a_6} \right\} \rightarrow \{d_2\}, \\ r_{x_2}^1 &: \left\{ \frac{0.84}{a_1}, \frac{0}{a_3}, \frac{0.93}{a_6} \right\} \rightarrow \{d_2\}, \\ r_{x_3}^1 &: \left\{ \frac{0.97}{a_1}, \frac{0}{a_3}, \frac{0.40}{a_6} \right\} \rightarrow \{d_1\}. \end{aligned}$$

(2) For the new object x_4 , $U = \{x_1, x_2, x_3\}$, $U^* = \{x_1, x_2, x_3, x_4\}$. Since $f_C(x_i) (i = 1, 3)$ and $f_C^*(x_4)$ satisfy Theorem 2 and $f_C(x_2) \supset f_C^*(x_4)$ satisfies Theorem 3, which leads to the fact that $\vec{U} = \emptyset$ and $\vec{U} = \{x_2\}$. According to Equation (5),

$$\begin{aligned} PIS_{\vec{U}}(x_4, a_2) &= \emptyset; NIS_{\vec{U}}(x_4, a_2) = \emptyset; \\ PIS_{\vec{U}}(x_4, a_4) &= \emptyset; NIS_{\vec{U}}(x_4, a_4) = \emptyset; \\ PIS_{\vec{U}}(x_4, a_5) &= \emptyset; NIS_{\vec{U}}(x_4, a_5) = \{x_2\}. \end{aligned}$$

According to Equation (6), $PID_{\vec{U}}(x_4, a) = \frac{|PIS_{\vec{U}}(x_4, a)|}{|\vec{U}|}$ and $NID_{\vec{U}}(x_4, a) = \frac{|NIS_{\vec{U}}(x_4, a)|}{|\vec{U}|}$. Thus

$$\begin{aligned} PID_{\vec{U}}(x_4, a_2) &= 0; NID_{\vec{U}}(x_4, a_2) = 0; \\ PID_{\vec{U}}(x_4, a_4) &= 0; NID_{\vec{U}}(x_4, a_4) = 0; \\ PID_{\vec{U}}(x_4, a_5) &= 0; NID_{\vec{U}}(x_4, a_5) = \frac{1}{3}, \end{aligned}$$

then $C_2 \cup \{a_5\} = \{a_1, a_3, a_5, a_6\}$ is a granular consistent set of the updated fuzzy formal decision context. Moreover, we can obtain a granular reduct $C_2^* = \{a_5, a_6\}$ from C_2 . The updated granular rules are as follows:

$$\begin{aligned} r_{x_1}^2 &: \left\{ \frac{0.99}{a_5}, \frac{0}{a_6} \right\} \rightarrow \{d_2\}, \\ r_{x_2}^2 &: \left\{ \frac{0.65}{a_5}, \frac{0.93}{a_6} \right\} \rightarrow \{d_2\}, \\ r_{x_3}^2 &: \left\{ \frac{0.89}{a_5}, \frac{0.40}{a_6} \right\} \rightarrow \{d_1\}, \\ r_{x_4}^2 &: \left\{ \frac{0.87}{a_5}, \frac{0.49}{a_6} \right\} \rightarrow \{d_1\}. \end{aligned}$$

(3) For the new object x_5 , since $f_C(x_i) (i = 1, 2, 3, 4)$ and $f_C^*(x_5)$ satisfies Theorem 2, the granular reduct C_2^* is unchanged. The updated granular rules are as follows:

$$\begin{aligned} r_{x_1}^3 &: \left\{ \frac{0.99}{a_5}, \frac{0}{a_6} \right\} \rightarrow \{d_2\}, \\ r_{x_2}^3 &: \left\{ \frac{0.65}{a_5}, \frac{0.93}{a_6} \right\} \rightarrow \{d_2\}, \\ r_{x_3}^3 &: \left\{ \frac{0.89}{a_5}, \frac{0.40}{a_6} \right\} \rightarrow \{d_1\}, \\ r_{x_4}^3 &: \left\{ \frac{0.87}{a_5}, \frac{0.49}{a_6} \right\} \rightarrow \{d_1\}, \\ r_{x_5}^3 &: \left\{ \frac{0.86}{a_5}, \frac{0.52}{a_6} \right\} \rightarrow \{d_1\}. \end{aligned}$$

C. The fuzzy rule based classification model

In this subsection, we can design a classification model based on the fuzzy granular rule introduced. In other words, we need to find the most likely label from the learned fuzzy granular rule base for a waiting predicted object t . Thus, how to find such the most likely label is crucial. The traditional method is to calculate the similarity degree between it and each learned rule, and then the label of the rule corresponding to the maximum similarity degree will be assigned to t . Therefore, we need to introduce the similarity degree between fuzzy granular rules.

Fig. 1: The overall procedure of FRCM.

Definition 10: Let $S = (U, A, \tilde{I}, D, J)$ be a consistent fuzzy formal decision context, C be a granular reduct of S and $r_x, r_y \in R(S)$. Then, the similarity degree between them is defined as follows:

$$CS(r_x, r_y) = \frac{\sum_{i=1}^{|r_x|} \min(\tilde{I}(x, a_i), \tilde{I}(y, a_i))}{\sum_{i=1}^{|r_x|} \max(\tilde{I}(x, a_i), \tilde{I}(y, a_i))}, a_i \in C. \quad (7)$$

Example 4: Continue with Example 3. We can compute the similarity degree between r_{x_5} and $r_{x_i} (i = 1, \dots, 4)$ and classify x_5 as d_1 :

$$CS(r_{x_5}, r_{x_1}) = \frac{0.86 + 0}{0.99 + 0.52} = 0.56,$$

$$CS(r_{x_5}, r_{x_2}) = \frac{0.65 + 0.52}{0.86 + 0.93} = 0.65,$$

$$CS(r_{x_5}, r_{x_3}) = \frac{0.86 + 0.40}{0.89 + 0.52} = 0.89,$$

$$CS(r_{x_5}, r_{x_4}) = \frac{0.86 + 0.49}{0.87 + 0.52} = 0.97.$$

We show the overall framework of FRCM in Fig. 1, in which the different shapes are employed to represent attributes and slivers are used to denote fuzzy granular rules, respectively. As vividly depicted, in the first stage, FRCM calculates a granular reduct for the initial dataset, and further initializes fuzzy granular rule base based on the granular reduct. In the second stage, given an incremental object x , FRCM goes into the incremental learning process to update the granular reduct to generate a new fuzzy granular rule set. After the incremental process is finished, it calculates the similarity degree

between the unlabeled object and each obtained fuzzy granular rules and the label of the granular rule corresponding to the maximum similarity degree will be assigned to the unlabeled object. Detailed description can be found in Algorithm 3.

Algorithm 3 Prediction algorithm based on FRCM

Input: Testing set T and the granular rule set $R(S)$.

Output: The prediction accuracy Acc.

```

1: Set Count = 0;
2: For each  $t$  in  $T$ 
3:   Compute the similarity degree between the object  $t$ 
   and each fuzzy granular rule in  $R(S)$  by Definition 10;
4:   Assign a label to  $t$ ;
5:   If predict correctly
6:     Count = Count + 1;
7:   End
8: End
9: Return Acc=Count/ $|T|$ .
```

D. Time complexity analysis

In this subsection, we analysis the time complexity of the proposed algorithms in this paper.

In Algorithm 1, Step 1 is the initialization process and Steps 2-8 aim at selecting a candidate attribute set C_+ , which makes $C \cup C_+$ become a granular consistent set of the updated fuzzy formal decision context. In order to get a minimum candidate attribute set, the attribute with the strongest identification ability will be selected one by one. Since $C \cup C_+$ is not always a granular reduct, Step3 9-13 and Steps 14-18 will filter out the extra attributes in the newly added attribute set and the original attribute set, respectively. In summary, the time complexity of Algorithm 1 in the worst case is $O(|U_*||A| + |U||C|)$.

In Algorithm 2, Steps 1-3 are the initialization process and Steps 4-12 update the granular rules according to Theorem 9. Since it recalls Algorithm 1, it is easy to verify that its time complexity is $O(|U_*||A| + |U||C| + |R(S)|)$.

Algorithm 3 predicts label for each unlabeled object by calculating similarity with the time complexity of $O(|R(S)||T|)$.

IV. EXPERIMENTAL RESULTS

In this section, numerical experiments are given to verify the effectiveness of the proposed classification method. In the experiments, we first compare FRCM with two types of classification algorithms: one is the rule-based classification methods and the other is the non-rule-based classification methods. Specifically, the former includes Decision Table [58], FURIA [59], JRip [60], OneR [61], PART [62], Ridor [63], REPSO [64] and SGERD [65], while the latter covers IBK [66], FLR [67], BayesNet, LWL [68], J48 [69], Random Forest [70], REPTree, SMO and IFKNN [71]. Besides, considering that the fuzzy granular rule based classification can also be implemented directly without granular reduct (the simplified FRCM named S-FRCM) for a given dataset, which will save much more time than FRCM, we also make a comparison of FRCM and S-FRCM to evaluate the necessity of granular

reduct. We then evaluate the dynamic learning ability of FRCM on classification task. Finally, we make a comparison between the incremental granular reduct algorithm proposed in this paper and the discernibility attribute matrix based method [29].

In the experiment, ten fold cross-validation was performed to obtain the average accuracy, and the training set will be used for granular reduct computing to obtain the granular rules, while the testing set will be used to verify the classification ability of the obtained granular rules. Note that, FRCM and S-FRCM were trained by MATLAB 2016b, while other classification algorithms (except for REPSO, SGERD and IFKNN were implemented in KEEL) were trained by Weka 3.8.4 and all the parameters of models were selected by default. For fairness, they were run on the same computer with Intel Core i7-4720HQ@2.60GHz CPU and 16GB main memory.

A. The datasets

For experiment, twenty-four public datasets were downloaded from some popular machine learning databases, such as UCI Machine Learning Repository and Weka. Please see Table III for details. Most of these datasets are numeric, except for individual attributes which are character type. For the non-numeric attributes, we replaced them with different numeric values. Note that this paper mainly concerns the fuzzy formal decision contexts, so some fuzzification methods were used to generate available datasets for experiments. Since all attribute values of Sonar are between 0 and 1, we do not do any processing on it and the remaining datasets were fuzzified by using the method in [72]. This method transforms the specific value into the membership degree. For instance, it changes attribute "Height" into three attributes: "High", "Middle" and "Low", and then converts the specific height into the membership values of "High", "Middle" and "Low". Therefore, the fuzzified datasets have more attributes. Besides, the inconsistent objects were eliminated to make these datasets consistent. For convenience, the first word of the datasets are used to represent the processed datasets, such as Lung for Lung Cancer. For more information, see Table IV.

B. Compare FRCM with rule-based classification methods

In this subsection, we compare FRCM with the rule-based classification methods, among which FURIA, REPSO, SGERD are fuzzy rule based algorithms while JRip is a propositional rule learner, OneR uses the minimum-error attribute for prediction and PART gets the rules by building a partial C4.5 decision tree, as well as Ridor focuses on the ripple down rules. The evaluation results are remarked in Table V and the best results are shown in bold.

As listed in Table V, FRCM reduced all of the selected datasets and showed better classification performance in 16 of them. Compared with S-FRCM, FRCM showed better performance on most of the datasets except for Iris, Vehicle, OBS, Segment, Mushroom and Pendigits, which exactly shows that appropriate granular reduct can indeed improve the classification performance of the proposed model. This may

TABLE III: The Statistics of the Chosen Datasets

ID	Dataset	#Object	#Attribute	#Class	Attribute Type
1	Lung Cancer	32	56	2	Categorical, Integer
2	Zoo	101	17	7	Categorical, Integer
3	Breast Cancer Coimbra	116	9	2	Integer, Real
4	Lymph	148	18	4	Categorical, Integer
5	Iris	150	4	3	Real
6	Hepatitis	155	19	2	Categorical, Integer, Real
7	Wine	178	13	2	Integer, Real
8	Sonar	208	60	2	Real
9	Glass	214	9	6	Real
10	Cleveland	303	13	5	Categorical, Integer, Real
11	Ecoli	336	7	8	Real
12	Ionosphere	351	34	2	Categorical, Real
13	Dermatology	366	34	6	Categorical, Integer
14	Monk	601	6	2	Categorical
15	Credit-a	690	15	2	Categorical, Integer, Real
16	Vehicle	846	18	4	Integer
17	Unbalanced	856	32	2	Categorical, Integer, Real
18	Vowel	990	14	11	Categorical, Real
19	OBS-Network	1075	22	4	Integer
20	Segment	2310	19	7	Integer, Real
21	Sick	3772	27	2	Categorical, Integer, Real
22	Page Blocks Classification	5473	10	5	Integer, Real
23	Mushroom	8124	22	2	Categorical, Integer
24	Pendigits	10992	16	10	Integer

TABLE V: The Average Prediction Accuracy (Mean±Standard Deviation %) of FRCM and Several Rule-based Classification Methods

Dataset	FRCM	S-FRCM	Decision Table	FURIA	JRip	OneR	PART	Ridor	REPSO	SGERD
Lung	97.50±7.91	60.83±30.94	83.33±22.22	66.67±20.79	80.83±27.23	64.17±28.88	79.17±18.94	70.00±23.31	71.67±27.27	70.83±23.98
Zoo	99.00±3.16	97.09±4.69	88.18±11.22	90.09±10.54	84.27±14.03	71.45±17.38	93.18±7.93	90.18±9.23	87.19±10.47	86.00±9.66
Breast	81.06±3.27	62.80±14.80	73.79±11.14	75.83±12.48	63.48±15.98	67.58±9.17	67.73±16.74	66.67±16.85	57.20±12.96	63.33±15.81
Lymph	85.81±10.63	77.00±10.77	74.95±11.57	79.00±13.71	80.95±12.05	76.38±12.87	77.52±6.65	74.24±15.64	65.24±14.12	72.67±12.75
Iris	100.00±0.00	100.00±0.00	100.00±0.00	99.09±2.87	98.09±4.03	96.36±8.78	100.00±0.00	97.18±4.54	98.00±4.03	99.00±3.16
Hepatitis	87.08±4.33	79.42±10.82	83.08±10.17	78.08±10.35	79.33±8.12	80.50±10.50	80.04±11.12	80.58±9.45	83.10±12.66	82.54±11.15
Wine	99.44±1.76	98.30±2.74	96.01±4.71	93.79±6.71	96.08±5.28	83.76±12.07	92.68±5.93	93.20±7.46	72.29±14.32	89.87±10.46
Sonar	88.02±4.56	83.67±8.45	65.95±7.95	79.40±8.87	76.00±7.65	62.52±9.71	73.67±11.77	73.12± 8.08	54.03±12.34	68.89±13.56
Glass	75.45±9.21	72.32±10.02	64.89±10.54	66.75±9.13	59.13±8.46	53.85±10.17	70.43±8.22	64.81±10.64	44.50±13.32	51.06±11.55
Cleveland	58.09±5.30	50.18±7.05	58.77±6.58	56.81±5.06	55.13±6.66	53.47±6.03	49.16± 6.14	55.82±8.36	53.13±5.37	46.71±5.87
Ecoli	72.25±3.03	71.94±4.75	68.61±8.81	75.91±4.65	71.66±6.67	63.75±5.26	73.89±7.25	72.86±3.07	71.68±6.64	72.26±7.22
Ionosphere	92.02±2.24	91.46±2.97	90.61±3.51	90.60±4.48	89.17±4.83	84.61±3.87	87.75±4.03	87.75±5.54	63.93±8.23	79.20±7.86
Dermatology	99.50±1.58	97.97±4.84	94.39±7.28	95.45±5.99	97.97± 3.52	56.61±9.72	96.95±4.91	93.76±11.56	35.76±5.19	89.82±9.24
Monk	100.00±0.00	94.14±4.33	75.04±8.63	97.83±4.52	91.87±10.00	75.04±8.63	96.68±4.07	92.35±6.95	75.04±8.63	64.43±8.68
Credit-a	83.47±2.91	83.48±5.47	84.93±1.95	85.80±4.03	86.38±2.99	85.51±2.46	84.06±3.20	84.06±4.48	80.80±4.20	85.51± 2.46
Vehicle	63.56±5.85	68.42±4.26	59.42±4.81	67.94±6.43	64.75±6.42	47.24±6.21	69.36±2.79	68.52±5.55	37.49±6.04	53.34±7.30
Unbalanced	98.60±1.43	98.13±1.36	98.60±1.44	95.38±10.19	98.60±1.44	98.48±1.36	98.13±1.92	98.60±1.44	98.60±1.44	97.54±2.68
Vowel	98.79±1.24	97.98±1.84	52.73±2.92	77.37±4.86	62.83±4.61	20.51±4.26	71.72±3.08	65.86±5.42	19.90±10.26	34.14±4.82
OBS	100.00±0.00	100.00±0.00	99.72±0.89	100.00±0.00	99.91±0.29	90.14±2.49	100.00±0.00	99.43±1.79	53.10±2.92	75.38± 6.49
Segment	96.27±1.08	96.45±0.95	85.96±2.43	96.10±0.95	94.58±1.40	62.29±4.03	95.75±1.25	94.54±0.72	60.71±15.09	81.54±4.54
Sick	96.76±1.40	96.05±1.00	92.64±14.42	98.43±0.76	97.95±0.69	96.22±1.06	97.93±1.01	98.18±0.98	92.85±1.77	94.10±1.31
Page	97.29±0.58	97.22±0.60	96.41±1.24	97.54±0.76	97.44±0.92	94.55±0.89	97.26±0.81	97.22±1.01	90.52±1.00	93.37±1.12
Mushroom	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	95.72±0.57	100.00±0.00	100±0.00	94.58±1.13	90.68±1.80
Pendigits	98.05±0.39	99.19±0.24	74.58±1.35	97.62±0.47	95.77±0.57	38.83±1.91	96.54±0.63	95.27±0.92	69.14±6.64	56.09±1.44
Ave. Acc.	90.34	86.42	81.77	85.89	84.26	71.65	85.40	83.92	67.93	74.93

be due to the fact that attribute reduction can eliminate the noise data.

When comparing with the remaining eight rule-based classification methods, FRCM can achieve state-of-the-art results in all of the chosen datasets except for Cleveland, Ecoli, Credit-a, Vehicle, Sick, Page and Mushroom. At the same time, its average classification accuracy is 90.34%, which is 3.92% higher than the second ranked S-FRCM and 22.41% higher than the last ranked REPSO. Besides, it seems that FRCM has better overall performance and more stable than the

selected classification algorithms. All of these results vividly illustrate that appropriate granular reduct can greatly improve the classification ability of the proposed model, and the fuzzy granular rule is a good kind of classification rules.

C. Compare FRCM with non-rule-based classification methods

In this subsection, we compare FRCM with several popular non-rule-based classification methods, that is three kinds of trees methods, two popular nearest neighbor algorithms, one

TABLE VI: The Average Prediction Accuracy (Mean±Standard Deviation%) of FRCM and Several Non-rule-based Classification Methods

Dataset	FRCM	IBK	FLR	BayesNet	LWL	J48	RandomForest	REPTree	SMO	IFKNN
Lung	97.50±7.91	67.50±31.53	88.33±15.31	77.50±30.944	76.67±26.29	68.33±33.74	73.33±25.09	75.00±23.89	70.83±28.66	71.67±27.27
Zoo	99.00±3.16	96.09±5.05	90.18±9.22	96.09±6.91	86.18±6.84	92.18±7.59	95.09±5.18	87.00±14.18	93.18±7.92	89.18±7.15
Breast	81.06±3.27	74.77±13.50	52.88±15.07	63.26±18.98	72.88±9.23	61.67±17.05	72.27±15.90	59.47±13.18	68.56±18.10	69.85±9.06
Lymph	85.81±10.63	77.00±11.71	41.52±15.05	76.90±11.94	76.43±12.49	76.95±10.01	83.05±10.72	76.24±11.57	81.67±9.18	77.67±11.60
Iris	100.00±0.00	100.00±0.00	99.00±3.16	100.00±0.00	94.36±8.87	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00	100.00±0.00
Hepatitis	87.08±4.33	80.71±7.24	84.36±8.39	81.29±9.44	83.13±11.08	76.79±11.16	82.50±7.18	79.29±8.18	80.67±8.196	83.08±9.13
Wine	99.44±1.76	98.33±3.74	74.51±14.60	95.56±5.73	88.27±7.99	93.27±4.36	96.64±5.98	93.79±5.59	96.63±3.90	97.78±5.36
Sonar	88.02±4.56	86.10±8.77	60.57±9.75	78.93±9.19	75.62±10.72	74.62±9.88	81.83±9.65	70.76±8.01	79.31±7.51	82.29±11.18
Glass	75.45±9.21	70.35±11.05	61.02±9.51	64.87±6.94	53.61±11.95	70.02±10.19	78.91±10.00	55.39±12.35	70.87±7.33	69.52±7.37
Cleveland	58.09±5.30	54.46±6.64	47.98±14.22	52.45±5.81	57.12±4.09	50.80±8.47	56.77±4.41	54.77±8.94	54.49±7.48	57.75±6.39
Ecoli	72.25±3.03	71.62±5.52	62.84±7.50	75.89±4.77	70.45±6.32	74.37±5.87	78.35±5.79	71.95±6.67	77.43±4.35	79.85±6.70
Ionosphere	92.02±2.24	91.75±4.07	91.17±6.91	89.75±4.04	87.16±7.30	86.31±7.96	91.75±3.38	90.02±5.27	91.18±3.35	90.89±3.49
Dermatology	99.50±1.58	97.95±3.54	98.45±2.50	99.50±1.58	88.68±11.19	96.95±4.91	99.50±1.58	95.92±5.74	99.50±1.58	98.97±2.16
Monk	100.00±0.00	89.391±5.56	66.67±11.12	74.33±8.26	75.04±8.63	98.89±3.51	98.89±2.49	92.71±8.27	75.04±8.63	85.04±8.59
Credit-a	83.47±2.91	82.32±5.41	44.64±5.82	85.36±4.60	85.51±2.46	84.93±3.49	86.67±3.97	85.36±2.31	85.36±3.16	86.23±5.03
Vehicle	63.56±5.85	68.41±6.40	48.75±5.58	62.25±3.92	54.32±5.74	68.06±4.73	72.67±4.37	70.90±5.70	69.59±4.43	70.91±4.59
Unbalanced	98.60±1.43	97.55±1.49	98.60±1.44	98.60±1.44	98.48±1.46	98.60±1.44	98.36±1.76	98.60±1.44	98.36±1.58	98.60±1.44
Vowel	98.79±1.24	98.48±1.28	24.95±4.44	59.19±4.54	33.54±3.77	75.76±6.08	96.36±2.24	67.07±6.46	86.36±4.71	93.54±3.37
OBS	100.00±0.00	100.00±0.00	80.28±3.63	87.14±3.53	75.86±4.02	100.00±0.00	100.00±0.00	96.81±2.16	86.29±6.35	97.94±1.96
Segment	96.27±1.08	95.45±0.84	72.65±5.29	90.81±1.32	74.43±3.01	96.14±1.44	98.01±0.65	94.28±1.49	95.36±0.86	94.63±1.35
Sick	96.76±1.40	96.39±0.93	94.05±1.23	95.30±0.95	96.55±0.88	98.62±0.63	98.17±0.90	98.35±0.81	97.42±1.11	96.36±1.14
Page	97.29±0.58	97.00±0.64	92.17±1.36	91.10±1.49	93.92±1.06	97.22±0.88	98.06±0.63	96.91±0.96	95.93±0.99	97.13±0.49
Mushroom	100.00±0.00	100.00±0.00	66.37±10.38	93.54±0.69	92.63±0.56	99.98±0.07	100.00±0.00	100.00±0.00	99.93±0.13	100.00±0.00
Pendigits	98.05±0.39	99.19±0.31	25.09±3.36	86.44±0.95	69.84±2.24	96.13±0.64	99.04±0.34	94.86±0.70	98.35±0.45	99.08±0.38
Ave. Acc.	90.34	87.12	69.46	82.34	77.53	84.86	89.01	83.56	85.51	87.00

TABLE IV: The Generated Consistent Fuzzy Formal Decision Contexts

Dataset	#Object	#Training	#Testing	#Attribute	#Class
Lung	32	26	6	168	2
Zoo	101	81	20	48	7
Breast	116	93	23	27	2
Lymph	147	118	29	54	4
Iris	107	86	21	12	3
Hepatitis	155	124	31	57	2
Wine	178	143	35	39	2
Sonar	208	167	41	60	2
Glass	213	171	42	27	6
Cleveland	303	243	60	39	5
Ecoli	328	263	65	21	8
Ionosphere	351	281	70	99	2
Dermatology	193	155	38	102	6
Monk	273	219	54	18	2
Credit-a	690	483	207	42	2
Vehicle	845	676	169	54	4
Unbalanced	856	685	171	96	2
Vowel	990	792	198	39	11
OBS	1064	444	111	60	4
Segment	2307	1846	461	54	7
Sick	3764	3012	752	81	2
Page	5399	4320	1079	30	5
Mushroom	8124	5687	2437	60	2
Pendigits	10992	8794	2198	48	10

Bayes method, LWL and SMO. In the experiments, we set $k = 1$ for IBK and chose the Radial Kernel function for SMO.

The evaluation results are reported in Table VI and the best ones are marked in bold. From these results, we can observe that FRCM not only showed competitive performance on the dataset that already received good results by the other approaches, but also performed well on the datasets with poor performance of other algorithms, such as Breast. It achieved the best performance in 16 out of 24 datasets and

poorer performance in eight datasets, i.e. Glass, Ecoli, Credit-a, Vehicle, Segment, Sick, Page and Pendigits. Compared with these non-rule-based classification methods, it ranked No.1 with an average accuracy 90.34%, which exactly demonstrates that the granular reduct based rule classification method has better advantages in terms of classification task, and also shows the effectiveness of the algorithm proposed in this paper. All of these may can owed to the attribute reduction mechanism adopted in FRCM.

To analyze the performances of the compared algorithms systematically, we employed the Friedman test [73], which is a widely used statistical test, for comparing multiple methods over a number of datasets. Specifically, given k algorithms and N datasets, let r_i^j be the rank of the j th algorithm on the i th datasets, where average ranks are assigned in case of ties. Let $R_j = \frac{1}{N} \sum_{i=1}^N r_i^j$ represent the average rank for the j th algorithm. Under the null-hypothesis, that is, all the compared algorithms perform equivalently, the Friedman statistic F_F will be distributed according to the F-distribution with $k - 1$ numerator degrees of freedom and $(k - 1)(N - 1)$ denominator degrees of freedom,

$$F_F = \frac{(N-1)\chi_F^2}{N(k-1)-\chi_F^2},$$

where $\chi_F^2 = \frac{12N}{k(k+1)} \left[\sum_{j=1}^k R_j^2 - \frac{k(k+1)^2}{4} \right]$. Then, if F_F is larger than the critical value at significance level α , the null hypothesis will be rejected.

VII shows the Friedman's Anova Table ($N = 24, k = 19$) analysed by MATLAB, in which Chi-sq denotes the Friedman statistics F_F . In order to further analyze the performance between FRCM and the selected algorithms, a certain post-

TABLE VII: The Friedman's Anova Table

Source	SS	df	MS	Chi-sq	Pro>Chi-sq
Columns	5179.42	18	287.75	167.43	2.90647E-26
Error	8184.58	414	19.77		
Total	13364	455			

TABLE VIII: Post hoc comparison table for $\alpha = 0.05$

i	Algorithm	$z = (R_0 - R_i)/SE$	p	Holm
18	REPSO	7.310096	0	0.002778
17	OneR	6.912529	0	0.002941
16	SGERD	6.912529	0	0.003125
15	FLR	6.450839	0	0.003333
14	LWL	5.527458	0	0.003571
13	BayesNet	4.193687	0.000027	0.003846
12	Ridor	4.142388	0.000034	0.004167
11	Decision Table	4.116738	0.000038	0.004545
10	REPTree	4.001316	0.000063	0.005000
9	J48	3.347254	0.000816	0.005556
8	Jrip	3.334430	0.000855	0.006250
7	PART	3.013811	0.002580	0.007143
6	S-FRCM	2.590595	0.009581	0.008333
5	SMO	2.552121	0.010707	0.010000
4	FURIA	2.500822	0.012391	0.012500
3	IBK	2.411049	0.015907	0.016667
2	IFKNN	1.615916	0.106112	0.025000
1	Random Forest	0.218020	0.827413	0.050000

hoc test, namely Holm's method [74] needs to be done. Let $p_1, p_2, \dots, p_{k-1}$ be the ordered p-values, with $p_1 \leq p_2 \leq \dots \leq p_{k-1}$, and $H_1, H_2, \dots, H_{k-1}$ be the corresponding hypotheses. The Holm procedure will reject $H_1, H_2, \dots, H_{i-1}$ if i is the smallest integer such that $p_i > \alpha/(k-1)$. The comparison results are shown in VIII, and Holm's procedure rejects those hypotheses that have an unadjusted p-value ≤ 0.008333 .

In addition, we also employed the Wilcoxon signed-rank test [75] for pairwise comparison between those algorithms. Let d_i be the difference between the performance of the two methods on the i th dataset. The differences are ranked according to their absolute values: the smallest d_i is assigned to the first rank. In the case of ties, average ranks are assigned. Let R^+ be the sum of the ranks for the datasets, on which the proposed method outperforms the compared methods, and it is defined as

$$R^+ = \sum_{d_i > 0} \text{rank}(d_i) + \frac{1}{2} \sum_{d_i = 0} \text{rank}(d_i)$$

and let R^- be the sum of the ranks for the datasets on which the compared methods outperform the proposed method. Then, according to the Wilcoxon's test, if $\min(R^+, R^-)$ is equal to or less than the critical values (81), the null hypothesis of equal performance is rejected. The results obtained by the Wilcoxon test for algorithm FRCM are listed in Tables IX.

D. Performance evaluation of the incremental learning ability

The main objective of incremental learning is to assess the ability of a classification model to learn knowledge from new data. As a classification model with an updating mechanism, FRCM may have better incremental learning ability. Therefore, in what follows, we mainly evaluate its ability to deal with new data.

TABLE IX: Results obtained by the Wilcoxon test for algorithm FRCM

VS	R^+	R^-	Exact P-value	Asymptotic P-value
S-FRCM	238.5	37.5	0.001369	0.002127
Decision Table	285.5	14.5	1.4723E-5	0.000102
FURIA	246.5	53.5	0.004573	0.005581
Jrip	250.0	26.0	2.534E-4	0.000622
OneR	297.0	3.0	5.96E-7	0.000025
PART	238.5	37.5	0.001369	0.002127
Ridor	249.0	27.0	2.98E-4	0.000696
REPSO	276.0	0.0	2.384E-7	0.000025
SGERD	295.0	5.0	1.192E-6	0.000032
IBK	246.5	29.5	4.416E-4	0.000916
FLR	300.0	0.0	1.192E-7	0.000017
BayesNet	283.5	16.5	2.241E-5	0.000129
LWL	295.0	5.0	1.192E-6	0.000032
J48	257.5	42.5	0.001319	0.002031
Random Forest	181.0	119.0	≥ 0.2	0.368120
REPTree	269.5	30.5	2.587E-4	0.000607
SMO	245.5	54.5	0.005070	0.006091
IFKNN	240.5	59.5	0.008317	0.009322

Incremental learning needs three kinds of datasets: initial dataset, incremental dataset and testing dataset. To generate these three kinds of datasets, we divided each dataset into 8 equal sized batches named Batches 1-8. During training, Batch 1 was taken as the initial set and Batches 2-7 were served as incremental sets. In the incremental learning, Batches 2-7 were added to the initial set in turns and Batch 8 was regarded as the testing set to verify the ability of incremental learning. For instance, in the second incremental process, Batches 1-2 are the initial set, Batch 3 is the incremental dataset, and Batch 8 is the testing dataset. Finally, we can complete six incremental learning processes. Note that the consistency of the datasets enables each object to induce a granular rule, therefore, the number of the granular rules obtained in the incremental process is equal to the size of the added batches. We use the accuracy line charts in Fig. 2 to record the incremental learning results of each dataset.

It can be observed from Fig. 2 that the prediction accuracy is constantly improved on most of the datasets with the incremental process goes on, and achieves the best classification performance in the final incremental process, which implies that the updating mechanism proposed in this paper is indeed effective. Besides, the optimal result is achieved in the fifth incremental process for Lymph, Monk, OBS, Segment and Page, while the fourth incremental process for Dermatology and Mushroom. Since the newly added objects will inevitably bring noise and then affect the discrimination ability of the proposed model, these results are reasonable.

We also list the running time of each incremental learning process in Table X. Obviously, as the learning process increases, more time is needed and the running time is greatly affected by the objects. Thus, the training is faster in the earlier process and will become slower in the later process, which leads to the fact that it only takes less than 0.11s to train Lung with 32 objects, but spends 325s to train Pendigits with 10992 objects. In addition, except for the number of objects, the complexity of dataset and the diversity of attributes can also affect the running time of FRCM. Thus, we also reported the average running time of the 24 datasets, which is the average running

Fig. 2: The evaluation results for incremental learning

time of six incremental processes. Detailed data are 1.57s, 3.60s, 6.42s, 10.01s, 14.69s and 19.85s, respectively, and they can be found at the end of Table X.

E. The comparison with other granular reduct algorithm

In this subsection, we mainly evaluate the running time of the incremental granular reduct algorithm proposed in this paper, and compare it with the discernibility matrix [29] based attribute reduction algorithm. For achieving this task, the first two objects of the whole dataset are used as the initial dataset to initialize the granular reduct. With the incremental process goes on, the remaining objects will be added to the initial dataset in turns to complete the updating of the granular reduct. Table XI lists the running time.

Obviously, due to the high complexity of calculating the discernibility matrix, the existing method is only effective for small datasets, but not for large datasets. Thus, the incremental algorithm proposed in this paper is more satisfactory in terms of computational efficiency.

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an update mechanism of fuzzy granular rules based on granular reduct, and further

designed a novel incremental learning algorithm FRCM for fuzzy datasets. In the whole process, the main difficulty lies in how to ensure the updated attribute set is a granular reduct rather than a granular consistent set. For solving this problem, we have presented some new notions, such as the identification ability of attributes and the similarity between fuzzy granular rules. Based on these results, the algorithms related to FRCM have also been developed to update fuzzy granular rules. Finally, a comparison with other popular classification algorithms has been made to demonstrate that FRCM can achieve better classification performance.

In fact, as an incremental learning algorithm, the performance of FRCM is unavoidable to be challenged by large-scale datasets. Therefore, a faster update method, such as batch updating needs to be considered. In addition, FRCM only involves the consistent formal decision contexts and cannot deal with inconsistent ones. Thus, an updating framework for inconsistent formal decision contexts is also worth exploring. At the same time, it is also a meaningful problem to apply FRCM for unsupervised learning due to the inaccuracy of the labels in real datasets. Therefore, we will focus on solving these problems in the future work.

TABLE X: The Running Time of the Incremental Learning Process(s)

Dataset	Pro. 1	Pro. 2	Pro. 3	Pro. 4	Pro. 5	Pro. 6
Lung	0.04	0.05	0.08	0.09	0.10	0.11
Zoo	0.04	0.05	0.06	0.06	0.06	0.07
Breast	0.01	0.02	0.02	0.03	0.03	0.04
Lymph	0.05	0.08	0.10	0.11	0.14	0.18
Iris	0.02	0.02	0.02	0.03	0.04	0.04
Hepatitis	0.05	0.06	0.08	0.11	0.13	0.16
Wine	0.03	0.03	0.03	0.04	0.05	0.07
Sonar	0.04	0.05	0.06	0.08	0.09	0.11
Glass	0.03	0.04	0.05	0.07	0.09	0.11
Cleveland	0.07	0.12	0.14	0.18	0.23	0.31
Ecoli	0.03	0.04	0.06	0.08	0.10	0.13
Ionosphere	0.09	0.13	0.16	0.19	0.23	0.28
Dermatology	0.10	0.13	0.14	0.16	0.19	0.20
Monk	0.04	0.04	0.06	0.05	0.07	0.07
Credit-a	0.11	0.18	0.26	0.37	0.52	0.69
Vehicle	0.30	0.62	1.05	1.55	2.12	2.74
Unbalanced	0.08	0.11	0.15	0.19	0.24	0.29
Vowel	0.20	0.42	0.67	1.03	1.45	1.95
OBS	0.09	0.11	0.11	0.11	0.11	0.11
Segment	0.94	1.95	3.45	5.17	7.15	9.33
Sick	2.49	6.10	10.13	15.39	23.32	30.59
Page	1.53	3.45	6.06	8.81	13.11	17.85
Mushroom	6.53	15.23	27.53	43.98	63.92	85.68
Pendigits	24.86	57.30	103.65	162.44	238.97	325.24
Ave.	1.57	3.60	6.42	10.01	14.69	19.85

TABLE XI: Running Time of the Reduction Calculation(s)

Dataset	Our work	Reference [29]
Lung	0.14	0.01
Zoo	0.10	0.05
Breast	0.08	0.33
Lymph	0.22	0.68
Iris	0.05	0.14
Hepatitis	0.16	1.65
Wine	0.16	2.34
Sonar	0.21	2.97
Glass	0.22	0.94
Cleveland	0.39	6.94
Ecoli	0.32	5.48
Ionosphere	0.60	37.83
Dermatology	0.27	0.53
Monk	0.10	14.24
Credit-a	2.30	778.81
Vehicle	3.72	427.02
Unbalanced	0.61	707.70
Vowel	2.84	75.98
OBS	0.11	133.40
Segment	13.16	789.50
Sick	38.21	>1h
Page	32.66	>1h
Mushroom	135.23	>1h
Pendigits	538.99	>1h

ACKNOWLEDGMENT

This work was supported by the National Key R&D Program of China (2020YFB1707801) and the National Natural Science Foundation of China (No. 11971211).

REFERENCES

- [1] B. Liu, W. Hsu, and Y. M. Ma, "Integrating classification and association rule mining," in *Proceedings of KDD*, vol. 1711, pp. 80–86, Aug. 1998.
- [2] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 13, no. 4, pp. 428–435, Aug. 2005.
- [3] P. P. Angelov and X. Zhou, "Evolving fuzzy-rule-based classifiers from data streams," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1462–1475, Apr. 2009.
- [4] Y. X. Chen and J. Wang, "Support vector learning for fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 11, no. 6, pp. 716–728, Dec. 2004.
- [5] H. Ishibuchi, K. Nozaki, H. Tanaka, *et al.*, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets and Systems*, vol. 52, no. 1, pp. 21–32, Nov. 1992.
- [6] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Efficient fuzzy partition of pattern space for classification problems," *Fuzzy Sets and Systems*, vol. 59, no. 3, pp. 295–304, Nov. 1993.
- [7] C. T. Sun, "Rule-base structure identification in an adaptive-network-based fuzzy inference system," *IEEE Transactions on Fuzzy Systems*, vol. 2, no. 1, pp. 64–73, Feb. 1994.
- [8] K. Nozaki, H. Ishibuchi, and H. Tanaka, "Adaptive fuzzy rule-based classification systems," *IEEE Transactions on Fuzzy Systems*, vol. 4, no. 3, pp. 238–250, Aug. 1996.
- [9] X. Z. Wang, L. C. Dong, and J. H. Yan, "Maximum ambiguity-based sample selection in fuzzy decision tree induction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 8, pp. 1491–1505, Aug. 2012.
- [10] X. Z. Wang, H. J. Xing, Y. Li, *et al.*, "A study on relationship between generalization abilities and fuzziness of base classifiers in ensemble learning," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 5, pp. 1638–1654, Oct. 2015.
- [11] P. Honko, "Granular computing for relational data classification," *Journal of Intelligent Information Systems*, vol. 41, no. 2, pp. 187–210, Mar. 2013.
- [12] R. Wille, "Restructuring lattice theory: an approach based on hierarchies of concepts," *Order Sets*, pp. 445–470, Sep. 1982.
- [13] S. O. Kuznetsov, "Complexity of learning in concept lattices from positive and negative examples," *Discrete Applied Mathematics*, vol. 142, no. 1–3, pp. 111–125, Aug. 2004.
- [14] J. M. Ma, M. J. Cai, and C. J. Zou, "Concept acquisition approach of object-oriented concept lattices," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 1, pp. 123–134, Aug. 2017.
- [15] W. X. Zhang, J. M. Ma, and S. Q. Fan, "Variable threshold concept lattices," *Information Sciences*, vol. 177, no. 22, pp. 4883–4892, Nov. 2007.
- [16] R. Belohlavek, V. Sklenar, and J. Zaczal, "Crisply generated fuzzy concepts," in *Proceedings of the Third International Conference on Formal Concept Analysis*, pp. 269–284, Feb. 2005.
- [17] J. H. Li, C. C. Huang, J. J. Qi, *et al.*, "Three-way cognitive concept learning via multi-granularity," *Information Sciences*, vol. 378, pp. 244–263, Feb. 2017.
- [18] A. Burusco and R. Fuentes-gonzalez, "The study of the L-fuzzy concept lattice," *Mathware and Soft Computing*, vol. 1, no. 3, pp. 209–218, Dec. 1994.
- [19] Y. L. Mi, Y. Shi, J. H. Li, *et al.*, "Fuzzy-based concept learning method: Exploiting data with fuzzy conceptual clustering," *IEEE Transactions on Cybernetics*, 2020, to be published, doi: 10.1109/TCYB.2020.2980794.
- [20] H. Yu, J. Lu, and G. Q. Zhang, "Topology learning-based fuzzy random neural network for streaming data regression," *IEEE Transactions on Fuzzy Systems*, vol. PP, no. 99, Nov. 2020, doi:10.1007/3-540-46027-6.
- [21] Y. L. Song, G. Q. Zhang, H. Y. Lu, *et al.*, "A self-adaptive fuzzy network for prediction in non-stationary environments," in *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Jul. 2018, pp. 1–8.
- [22] E. A. Soares, H. D. A. Camargo, S. J. Camargo, *et al.*, "Incremental gaussian granular fuzzy modeling applied to hurricane track forecasting," in *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Aug. 2018, pp. 1–8.
- [23] H. Yu, J. Lu, and G. Q. Zhang, "Online topology learning by a gaussian membership-based self-organizing incremental neural network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 3947–3961, Nov. 2019.
- [24] C. Garcia, D. Leite, and I. Škrjanc, "Incremental missing-data imputation for evolving fuzzy granular prediction," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 10, pp. 2348–2362, Aug. 2019.
- [25] Y. Shi, Y. L. Mi, J. H. Li, *et al.*, "Concept-cognitive learning model for incremental concept learning," *IEEE Transactions on Systems Man & Cybernetics Systems*, vol. 51, no. 2, pp. 809–821, Feb. 2021.
- [26] T. Nakashima, T. Sumitani, and A. Bargiela, "Incremental learning of fuzzy rule-based classifiers for large data sets," in *World Automation Congress*, Jun. 2012, pp. 1–5.

- [27] R. D. Baruah, P. P. Angelov, and J. Andreu, "Simpl-e-class: Simplified potential-free evolving fuzzy rule-based classifiers," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Oct. 2011, pp. 2249–2254.
- [28] W. Z. Wu, Y. Leung, and J. S. Mi, "Granular computing and knowledge reduction in formal contexts," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 10, pp. 1461–1474, Oct. 2009.
- [29] M. W. Shao, Y. Leung, X. Z. Wang, *et al.*, "Granular reducts of formal fuzzy contexts," *Knowledge-Based Systems*, vol. 114, pp. 156–166, Dec. 2016.
- [30] X. P. Kang, D. Y. Li, S. G. Wang, *et al.*, "Formal concept analysis based on fuzzy granularity base for different granulations," *Fuzzy Sets and Systems*, vol. 203, pp. 33–248, Sep. 2012.
- [31] M. W. Shao, M. M. Lv, K. W. Li, *et al.*, "The construction of attribute (object)-oriented multi-granularity concept lattices," *International Journal of Machine Learning and Cybernetics*, vol. 11, no. 5, pp. 1017–1032, May 2020.
- [32] W. H. Xu and W. T. Li, "Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 46, no. 2, pp. 366–379, May 2016.
- [33] W. H. Xu, J. Z. Pang, and S. Q. Luo, "A novel cognitive system model and approach to transformation of information granules," *International Journal of Approximate Reasoning*, vol. 55, no. 3, pp. 853–866, Mar. 2014.
- [34] L. A. Zadeh, "Fuzzy sets and information granularity," in *Advances in Fuzzy Set Theory and Application*, pp. 3–18, Amsterdam, the Netherlands: North-Holland Publishing, 1979.
- [35] Y. Y. Yao, "Three-way decisions and cognitive computing," *Cognitive Computation*, vol. 8, no. 4, pp. 543–554, Aug. 2016.
- [36] Y. Y. Yao, "Three-way decision and granular computing," *International Journal of Approximate Reasoning*, vol. 103, pp. 107–123, Dec. 2018.
- [37] J. H. Li, C. L. Mei, W. H. Xu, *et al.*, "Concept learning via granular computing: a cognitive viewpoint," *Information Sciences*, vol. 298, pp. 447–467, Mar. 2015.
- [38] J. J. Niu, C. C. Huang, J. H. Li, *et al.*, "Parallel computing techniques for concept-cognitive learning based on granular computing," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 11, pp. 1785–1805, Nov. 2018.
- [39] L. Wei, J. J. Qi, and W. X. Zhang, "Attribute reduction theory of concept lattice based on decision formal contexts," *Science in China Series F-Information Sciences*, vol. 51, no. 7, pp. 910–923, Jun. 2008.
- [40] W. X. Zhang, L. Wei, and J. J. Qi, "Attribute reduction theory and approach to concept lattices," *Science in China: Series F*, no. 6, pp. 713–726, Dec. 2005.
- [41] M. W. Shao, Y. Leung, and W. Z. Wu, "Rule acquisition and complexity reduction in formal decision contexts," *International Journal of Approximate Reasoning*, vol. 55, no. 1, pp. 259–274, Jan. 2014.
- [42] J. S. Mi, Y. Leung, and W. Z. Wu, "Approaches to attribute reduction in concept lattices induced by axialities," *Knowledge-Based Systems*, vol. 23, pp. 504–511, Aug. 2010.
- [43] J. H. Li, C. L. Mei, and Y. J. Lv, "A heuristic knowledge-reduction method for decision formal contexts," *Computers and Mathematics with Applications*, vol. 61, pp. 1096–1106, Feb. 2011.
- [44] D. G. Chen and S. Y. Zhao, "Local reduction of decision system with fuzzy rough sets," *Fuzzy Sets & Systems*, vol. 161, no. 13, pp. 1871–1883, Jul. 2010.
- [45] D. G. Chen, L. Zhang, S. Y. Zhao, *et al.*, "A novel algorithm for finding reducts with fuzzy rough set," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 2, pp. 385–389, Apr. 2012.
- [46] D. G. Chen, S. Y. Zhao, L. Zhang, *et al.*, "Sample pair selection for attribute reduction with rough set," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 11, pp. 2080–2093, Nov. 2012.
- [47] D. G. Chen, Q. H. Hu, and Y. P. Yang, "Parameterized attribute reduction with gaussian kernel based fuzzy rough sets," *Information Sciences*, vol. 181, no. 23, pp. 5169–5179, Dec. 2011.
- [48] C. Z. Wang, Y. Wang, M. W. Shao, *et al.*, "Fuzzy rough attribute reduction for categorical data," *IEEE Transactions on Fuzzy Systems*, vol. 28, no. 5, pp. 818–830, May 2020.
- [49] C. Z. Wang, Y. L. Qi, M. W. Shao, *et al.*, "A fitting model for feature selection with fuzzy rough sets," *IEEE Transactions on Fuzzy Systems*, vol. 25, no. 4, pp. 741–753, Jun. 2017.
- [50] Q. H. Hu, D. R. Yu, J. F. Liu, *et al.*, "Neighborhood rough set based heterogeneous feature subset selection," *Information Sciences*, vol. 178, no. 18, pp. 3577–3594, Sep. 2008.
- [51] Q. H. Hu, D. R. Yu, W. Pedrycz, *et al.*, "Kernelized fuzzy rough sets and their applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 11, pp. 1649–1667, Nov. 2011.
- [52] J. Y. Liang, F. Wang, C. G. Dang, *et al.*, "A group incremental approach to feature selection applying rough set technique," *IEEE Transactions on Knowledge & Data Engineering*, vol. 26, no. 2, pp. 294–308, Feb. 2013.
- [53] J. Y. Liang, J. R. Mi, W. Wei, *et al.*, "An accelerator for attribute reduction based on perspective of objects and attributes," *Knowledge-Based Systems*, vol. 44, pp. 90–100, May 2013.
- [54] D. Pei and J. S. Mi, "Attribute reduction in decision formal context based on homomorphism," *International Journal of Machine Learning & Cybernetics*, vol. 2, no. 4, pp. 289–293, Jul. 2011.
- [55] W. X. Zhang, L. Wei, and J. J. Qi, "Attribute reduction in concept lattice based on discernibility matrix," *Lecture Notes in Computer Science*, vol. 3642, pp. 157–165, Aug. 2005.
- [56] B. S. Yahia and A. Jaoua, "Discovering knowledge from fuzzy concept lattice," In: Kandel A., Last M., Bunke H. (eds) *Data Mining and Computational Intelligence. Studies in Fuzziness and Soft Computing*, vol. 68, pp. 167–190, Mar. 2001, Physica, Heidelberg.
- [57] D. Pei, M. Z. Li, and J. S. Mi, "Attribute reduction in fuzzy decision formal contexts," in *Proceedings of the 2011 International Conference on Machine Learning and Cybernetics*, pp. 204–208, Jul. 2011.
- [58] R. Kohavi, "The power of decision tables," in *Proceedings of the 8th European Conference on Machine Learning*, Apr. 1995, pp. 174–189.
- [59] J. Hühn and E. Hüllermeier, "FURIA: an algorithm for unordered fuzzy rule induction," *Data Mining & Knowledge Discovery*, vol. 19, no. 3, pp. 293–319, Apr. 2009.
- [60] W. W. Cohen, "Fast effective rule induction," in *Proceedings of the Twelfth International Conference on Machine Learning*, Jul. 1995, pp. 115–123.
- [61] O. Holte, "Very simple classification rules perform well on most commonly used datasets," *Machine Learning*, no. 11, pp. 63–90, Apr. 1993.
- [62] E. Frank and I. H. Witten, "Generating accurate rule sets without global optimization," in *Fifteenth International Conference on Machine Learning*, Jan. 1998, pp. 144–151.
- [63] V. Veeralakshmi, "Ripple down rule learner (ridor) classifier for iris dataset," *International Journal of Computer Science Engineering*, vol. 4, no. 3, pp. 79–85, May 2015.
- [64] L. Z. Yu, Z. W. Shi, and S. Y. Chen, "Rule discovery with particle swarm optimization," Jan 2004, pp. 291–296.
- [65] E. Mansoori, M. Zolghadri, and S. Katebi, "Sgerd: A steady-state genetic algorithm for extracting fuzzy classification rules from data," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 4, pp. 1061–1071, Sep. 2008.
- [66] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, pp. 37–66, Sep. 1991.
- [67] Y. G. Kaburlasos, I. N. Athanasiadis, and P. A. Mitkas, "Fuzzy lattice reasoning (flr) classifier and its application for ambient ozone estimation," *International Journal of Approximate Reasoning*, vol. 45, no. 1, pp. 152–188, Sep. 2007.
- [68] E. Frank, M. Hall, and B. Pfahringer, "Locally weighted naive bayes," in *19th Conference in Uncertainty in Artificial Intelligence*, Oct. 2012, pp. 249–256.
- [69] J. R. Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [70] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, Oct. 2001.
- [71] Y. Q. Wu, K. Ianiakiev, and V. V. Govindaraju, "Improved k-nearest neighbor classification," *Pattern Recognition*, vol. 35, no. 10, pp. 2311–2318, Oct. 2002.
- [72] S. K. Pal and P. Mitra, "Case generation using rough sets with fuzzy representation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 3, pp. 293–300, Apr. 2004.
- [73] J. Demiar and D. Schuurmans, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, vol. 7, no. 1, pp. 1–30, Jun. 2006.
- [74] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, Jan. 1979.
- [75] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, Dec. 1945.

Jiaojiao Niu received the M.Sc. degree in science from Kunming University of Science and Technology, Kunming, China, in 2018. Currently, she is pursuing the Ph.D. degree at North China Electric Power University, Beijing, China. Her research interests include machine learning, granular computing and formal concept analysis.

Degang Chen received the M.Sc. degree in science from Northeast Normal University, Changchun, Jilin, China, in 1994 and the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 2000. He was a postdoctoral fellow with Xi'an Jiaotong University, Xi'an, China, from 2000 to 2002 and with Tsinghua University, Tsinghua, China, from 2002 to 2004. Since 2006, he has been a professor with North China Electric Power University, Beijing, China. His research interests include fuzzy groups, fuzzy analysis, rough sets, and support vector machines.

chines.

Jinhai Li received the M.Sc. degree in science from the Guangxi University, Guangxi, China, in 2009, and the Ph.D. degree in science from the Xi'an Jiaotong University, Xi'an, China, in 2012. Currently, he is a professor with Kunming University of Science and Technology. His research interests include big data, cognitive computing, granular computing and formal concept analysis.

Hui Wang received the B.Sc. degree in computer science and the M.Sc. degree in artificial intelligence from Jilin University of China, Changchun, China, in 1985 and 1988, respectively, and the D. Phil. degree in informatics from the University of Ulster, Coleraine, U.K., in 1996. From 1988 to 1992, he was with Jilin University as a lecturer. He is currently a professor with School of Electronics, Electrical Engineering and Computer Science Queen's University Belfast. His research interests include machine learning, logics and reasoning, combinatorial data

analytics, and their applications in image, video, spectra and text analysis. He has more than 200 publications in these areas.