A dynamic rule-based classification model via granular computing

.

A dynamic rule-based classification model via granular computing

Jiaojiao Niu[a,*], Degang Chen[b], Jinhai Li[c], Hui Wang[d]

[a] *School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China*
[b] *School of Mathematics and Physics, North China Electric Power University, Beijing 102206, China*
[c] *Data Science Research Center, Kunming University of Science and Technology, Kunming 650500, Yunnan, China*
[d] *School of Electronics, Electrical Engineering and Computer Science Queen's University Belfast, Computer Science Building, 18 Malone Road, BT9 5BN Belfast, United Kingdom*

Abstract

As an effective tool for data representation and processing, granular computing has been incorporated into formal decision contexts for finding granular reducts to achieve the task of mining granular rules. However, the classification performance of granular rules has not been evaluated, and this type of method is not suitable for dynamic data. To solve this problem, the current study updates granular reducts and evaluates the obtained granular rules in terms of classification performance. Concretely, we first give a theoretical analysis of updating granular reducts and granular rules and then present a novel dynamic rule-based classification model (DRCM) based on the updating mechanism. Finally, we discuss the feasibility of the proposed model and compare it with several popular classification algorithms. The conducted experiments demonstrate that the granular reducts can improve the classification ability to a certain extent and that DRCM can achieve better classification performance on some consistent datasets.

Key words: granular computing, formal concept analysis, granular reduct, granular rule, dynamic learning, classification

1. Introduction

Rule-based classification approaches aim to achieve classification tasks by certain types of acquired rules, such as fuzzy rules [1–4] and formal concept analysis (FCA)-based [5] rules [6]. One of the biggest challenges faced by such classification methods is how to mine useful information from massive data to improve the classification ability of rules in terms of the speed and accuracy. Since a popular adopted method is to reduce useless attributes in target datasets, it is meaningful to seek a suitable method for attribute reduction.

Motivated by seeking an effective reduction algorithm, researchers have studied various attribute reduction methods [7–11] in different disciplines [5, 12–14]. Among them, rough set theory (RST) and formal concept analysis-based attribute reduction are the most explored methods, and many models have been developed to address a variety of different requirements. In the literature, Bazan et al. studied dynamic reducts [15, 16] and decision reducts [17] using the Boolean reasoning [18] approach to extract decision rules. Meanwhile, Skowron et al. [19] introduced a discernibility matrix and discernibility function for computing the reducts of an information system. Considering that an information system can be transformed into a formal context, the discernibility matrix method was further investigated in FCA for solving attribute reduction problems. For example, applicable discernibility matrices were also defined to

[☆]This work was supported by the National Key R&D Program of China (2020YFB1707801) and the National Natural Science Foundation of China (No. 12071131)

[*]Corresponding author

Email addresses: njjiao_92@163.com (Jiaojiao Niu), wscz973@163.com (Degang Chen), jhlixjtu@163.com (Jinhai Li), h.wang@qub.ac.uk (Hui Wang)

compute all of the reducts of a formal (decision) context [5, 20], i.e., the classical concept-based discernibility matrices [21, 22]. Different from RST, attribute reduction in FCA focuses on not only keeping certain consistencies [21] or decision-making ability unchanged, but also preserving the concept lattice structure unchanged [20]. However, since RST and FCA can be complementary to each other, many researchers have attempted to combine them to achieve better data analysis ability. Liu et al. [23] proposed a multi-step attribute reduction method and object reduction method for attribute-oriented and property-oriented concept lattices by justifying whether an attribute or an object is dispensable.

FCA takes the concept as the basic unit, and to achieve more concept knowledge discovery tasks, additional theories, such as three-way decision [24, 25] and granular computing [22, 26–28], were used to study multiple concepts. As a result, many types of notions [29, 30] have been proposed from various research aspects. A few studies have been performed on the classification problem from the perspective of concepts. For example, based on concept-cognitive learning [29, 31], Shi et al. [32] proposed a concurrent incremental learning technique by continuously accommodating newly added data to meet the requirements of classification tasks. Considering that the corresponding rules can be defined through the implications among sets of attributes, Kuznetsov [33] developed an algorithm for computing concepts to generate all possible concept-based classification rules (hypotheses). Although these FCA-based classification methods have shown good classification ability, there are some limitations. First, they lack the mechanism for learning rules dynamically, and the obtained rules must be retrained when new data are added. Second, these methods did not consider attribute reduction, which will lead to a large number of invalid rules, and affect the classification ability. Although compact rules can be obtained by the so-called pruning strategy, this approach still suffers from instability. Finally, because the FCA-based rules were generated based on concepts, it will take plenty of time to calculate the rules. Note that some parallel technologies [34] have been used to improve the calculation efficiency, but there is still room for improvement.

On the one hand, the incremental learning method can effectively avoid repeated training of the model when new objects are added. This approach refers to a learning system that can learn new knowledge from unseen samples and saves most of the knowledge that has been learned before. Because of its strong self-study ability, researchers have applied it to different disciplines to propose appropriate incremental algorithms [15, 16, 35–39, 41, 42], among which Bazan et al. [39] presented a rough set approach to vague concept approximation that realized a step towards approximate reasoning in complex dynamic systems. In addition, the recent concept-cognitive learning model (CCLM) [32] also adopted an incremental learning strategy. On the other hand, attribute reduction can reduce those attributes that are not helpful to the classification task, and then, a more compact rule base can be obtained from the reduced datasets.

In addition, granular computing (GrC), which explores the composition of parts, their interrelationships, and connections to the whole, can effectively improve the efficiency of knowledge discovery [26]. Since complex problems can be better analysed from the perspective of information granulation, some valuable work [22, 40, 43–47] has been accomplished over a short time. Wu et al. [22] integrated granular computing into the framework of FCA for granularity learning in the concept lattice. As a result, attribute reduction was implemented from a granularity viewpoint to propose granular reducts and generate granular rules. Different from other rules in FCA, the acquisition of granular rules is free of concepts, and the corresponding algorithm has linear time complexity. This finding motivates us to further achieve a classification task based on dynamic granular rules. In this paper, we propose a novel dynamic rule-based classification model named DRCM via GrC. The main contributions are as follows: (1) we analyse the mechanism of updating granular reducts and granular rules; (2) granular reducts are further used to design an incremental classification algorithm.

The remainder of this paper is organized as follows. Section 2 reviews some notions related to FCA. In Section 3, we first give an overall introduction to the proposed DRCM, and then explain the updating mechanism. In Section 4, numerical experiments are conducted on sixteen datasets selected from the UCI Machine Learning Repository to evaluate the classification performance of the proposed model. A summary and future work are given in Section 5.

2. Preliminary knowledge

In this section, we review some notions related to FCA, i.e., formal decision contexts, granular rules and granular reducts.

Table 1: A formal decision context $S = (U, C, I, D, J)$

U	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x
x_1	1	0	0	1	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	1	0	1	0	0
x_2	1	0	0	1	0	0	1	1	1	1	0	0	0	0	1	0	0	0	1	0	1	1	0	0
x_3	1	0	0	1	0	0	1	1	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0
x_4	0	1	1	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	1	1	0	0	1	0
x_5	1	0	0	1	0	0	1	1	1	1	0	0	0	0	1	0	0	0	1	0	1	1	0	0
x_6	1	0	0	1	0	0	0	1	1	1	0	0	0	0	1	0	0	0	1	1	0	1	0	0
x_7	0	1	1	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	1	1	0	0	1	0
x_8	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
x_9	1	0	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	1	0	1	1	0	0
x_{10}	1	0	0	1	1	0	0	1	1	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0

Definition 1. [5] A formal context is a triplet $F = (U, C, I)$, where $U = \{x_1, x_2, ..., x_n\}$ is a nonempty, finite set of objects called the universe of discourse, $C = \{a_1, a_2, ..., a_m\}$ is a nonempty, finite set of attributes, and $I \subseteq U \times C$ is a binary relation between U and C. Here, $I(x, a) = 1$ indicates that the object x has the attribute a while $I(x, a) = 0$ means the opposite.

Definition 2. [20, 33] A formal decision context is a quintuple $S = (U, C, I, D, J)$ in which (U, C, I) and (U, D, J) are called conditional formal context and decision formal context with $C \cap D = \emptyset$.

Generally, the concept lattices of the conditional formal context and decision formal context are called the conditional concept lattice and decision concept lattice, respectively.

Table 1 is a formal decision context $S = (U, C, I, D, J)$ randomly selected from the Zoo dataset (see Section 4 for details), where $U = \{x_i | i = 1, 2, ..., 10\}$ represents the animal names, $C = \{a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u\}$ denotes the attribute characteristics, such as legs and hairs, related to those objects and $D = \{v, w, x\}$ indicates the classes of animals. From Table 1, we know that the object x_1 has the attributes a, d, h, i, j, o and t, and it is classified into v.

FCA takes formal concepts as the basic unit of knowledge. In general, a concept (X, Y) is a tuple composed of an object set X (extent) and an attribute set Y (intent) such that $X^* = Y$ and $Y^* = X$, where "$*$" is a mapping between U and C as follows:

$$X^* = \{a \in C | \forall x \in X, I(x, a) = 1\}, X \subseteq U,$$
$$Y^* = \{x \in U | \forall a \in Y, I(x, a) = 1\}, Y \subseteq C.$$

A concept can be obtained by the basic operations of objects or attributes only. For example, (X, Y) can be depicted by $\vee_{x \in X}(x^{**}, x^*)$. In FCA, such a pair (x^{**}, x^*) is regarded as a granular concept and can be strictly defined below:

Definition 3. [5] Let $F = (U, C, I)$ be a formal context. For any $x \in U, a \in C$, (x^{**}, x^*) and (a^*, a^{**}) are object concepts and attribute concepts, respectively, called granular concepts of F.

Since this paper will discuss the dynamic updating of objects in a formal decision context, object granular concepts will be given more attention. Hereinafter, "$*_B$" is used to represent the mapping between the object set U and the attribute subset B. That is, for $X \subseteq U, Y \subseteq B$,

$$X^{*_B} = \{a \in B | \forall x \in X, I_B(x, a) = 1\}, X \subseteq U,$$
$$Y^{*_B} = \{x \in U | \forall a \in Y, I_B(x, a) = 1\}, Y \subseteq B,$$

where $I_B = I \cap (U \times B)$.

Proposition 1. Let $F = (U, C, I)$ be a formal context. For any $Y \subseteq B \subseteq C$ and $X \subseteq U$, we have

(1) $Y^{*_B} = Y^{*_C}$;
(2) $X^{*_B} \subseteq X^{*_C}$;
(3) $X^{*_C *_C} \subseteq X^{*_B *_B}$.

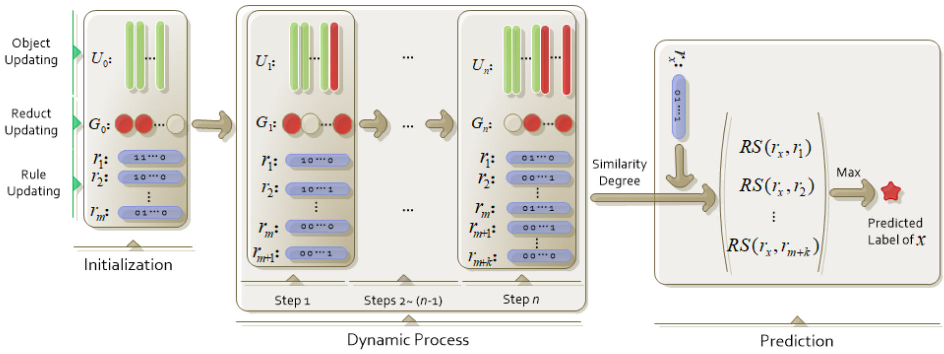

Figure 1: Illustration of overall procedure for DRCM.

The main purpose of analysing the formal decision context is to perform decision analysis, and rules are the main way to make decisions.

Definition 4. Let $S = (U, C, I, D, J)$ be a formal decision context. For any $x \in U$, $(x^{*_C *_C}, x^{*_C})$ and $(x^{*_D *_D}, x^{*_D})$ are granular concepts in the conditional concept lattice and decision concept lattice, respectively. Then $r_x : x^{*_C} \to x^{*_D}$ is called a granular rule of S, where x^{*_C} and x^{*_D} are called condition and conclusion (also a label of x) of r_x, respectively.

Due to massive data, an extracted granular rule base will inevitably contain a large number of useless rules for decision analysis. To refine the granular rule base, a useful way is to remove the useless attributes by means of granular reducts.

Definition 5. [22] Let $S = (U, C, I, D, J)$ be a formal decision context. S is said to be consistent if $x^{*_C *_C} \subseteq x^{*_D *_D}$ for any $x \in U$; otherwise, it is said to be inconsistent.

Definition 6. [22] Let $S = (U, C, I, D, J)$ be a formal decision context and $G \subseteq C$. If $x^{*_G *_G} \subseteq x^{*_D *_D}$ for any $x \in U$, then G is a granular consistent set of S and if there is no proper subset of G such that it is a granular consistent set of S, then G is referred to as a granular reduct of S.

Theorem 1. Let $S = (U, C, I, D, J)$ be a consistent formal decision context, G be a granular reduct of S and $E = C - G$, $K \subseteq E$. Then, for any $x \in U$, $x^{*_{G\cup K} *_{G\cup K}} \subseteq x^{*_D *_D}$.

Proof. It is obvious that $G \subseteq G \cup K \subseteq C$. According to Proposition 1 and Definition 7, $x^{*_{G\cup K} *_{G\cup K}} \subseteq x^{*_G *_G}$ and $x^{*_G *_G} \subseteq x^{*_D *_D}$. Thus, $x^{*_{G\cup K} *_{G\cup K}} \subseteq x^{*_D *_D}$ holds.

Theorem 1 shows that an attribute set that includes a granular reduct must be a granular consistent set of a consistent formal decision context.

Many rules can be extracted from a general formal decision context, but some of them might not be necessary in terms of classification. In addition, granular rules will become effective for achieving classification tasks when each object has only one label. For this purpose, we need to introduce the notion of a regular formal decision context.

Definition 7. [32] Let $S = (U, C, I, D, J)$ be a formal decision context. S is referred to as a regular formal decision context if for any $a, b \in D$, $a^* \cap b^* = \emptyset$.

3. The granular rule-based classification model

In this section, a dynamic updating framework named DRCM for computing granular reducts and granular rules is designed to achieve classification tasks. It includes four processes: (1) initialize granular reducts and granular rules; (2) perform granular reduct updating; (3) perform granular rule updating; and (4) generate predictions.

Figure 1 gives an overall procedure for DRCM in which “Max” indicates that we choose the maximal similarity degree (introduced in Section 3.3) to obtain the predicted label. Before the calculation, the input formal decision context will be divided into a training formal decision context (training set) and a testing formal decision context (testing set) with a given ratio. Then, the training set is used to learn the granular rule, while the testing set is employed to estimate the classification ability of the learned granular rules. During granular reduct and granular rule updating processes, the training set is further divided into an initial dataset and an incremental dataset, in which a new object is added into the initial dataset at a certain time, and then, we update the corresponding granular reduct.

The granular rules will be updated after the granular reduct updating is finished. In the process of prediction, for a testing instance, the similarity degree between it and each of the granular rules will be calculated, and the label of the granular rule with the maximal similarity degree will be assigned to the testing instance.

We consider that adding a new object will produce two states: the original formal decision context and the updated formal decision context. To distinguish them from each other, we use S and S^+ to represent two states of the formal decision context. In other words, for a formal decision context $S = (U, C, I, D, J)$, $S^+ = (U^+, C^+, I^+, D^+, J^+)$ is used to denote the updated version, where $U^+ = \{x_1, x_2, ..., x_n, x_{new}\}$, $C^+ = C$, $D^+ = D$ and $I^+ \subseteq U^+ \times C^+$, $J^+ \subseteq U^+ \times D^+$. To avoid confusion, we use "$\tilde{*}$" to represent the "$*$" mapping in the updated formal decision context S^+. In addition, all of the formal decision contexts to be discussed below are regular and consistent.

3.1. Granular reduct updating process

In this subsection, we mainly discuss the updating of granular reducts. The difficulty of this problem lies in judging when to update and how to update. As pointed out before, the granular reduct maintains the consistency of a formal decision context, which is closely related to the extents of the granular concepts. Therefore, we consider the extents of granular concepts as the beginning to discuss this problem.

Theorem 2. Let $S = (U, C, I, D, J)$ and $S^+ = (U^+, C^+, I^+, D^+, J^+)$ be two formal decision contexts and $K \subseteq C$. Then, for any $x \in U$, we have

$$x^{\tilde{*}_K \tilde{*}_K} = \begin{cases} x^{*_K *_K} \cup \{x_{new}\}, & \text{if } x^{*_K} \subseteq x_{new}^{\tilde{*}_K}, \\ x^{*_K *_K}, & \text{otherwise.} \end{cases}$$

Proof. It is obvious from the definitions of S and S^+.

Theorem 2 shows that adding a new object can make $x^{*_K *_K}$ increase, as does $x^{*_D *_D}$. Therefore, the changes in $x^{*_K *_K}$ and $x^{*_D *_D}$ will inevitably have an influence on the granular reduct G. Figure 2 presents a consistency analysis of G after adding a new object x_{new}, in which the results are caused by different combinations of the second and third columns. It is obvious that if the newly added object has the same label as one of the previous objects, it will certainly not affect the consistency of that object. Otherwise, it is not necessary. For convenience, for any two objects x and y, $Z(Z \neq x^*, Z \neq y^*)$ is used to denote their common attribute set. Then, we can summarize the following conclusions to facilitate subsequent analysis.

Theorem 3. Let $S = (U, C, I, D, J)$ and $S^+ = (U^+, C^+, I^+, D^+, J^+)$ be two formal decision contexts, and G be a granular reduct of S. G is a granular reduct of S^+ if any $x \in U$ satisfies one of the following conditions:

$$\begin{aligned} &(1) \quad x^{*_D} = x_{new}^{\tilde{*}_D}; \\ &(2) \quad x^{*_D} \neq x_{new}^{\tilde{*}_D}, x^{*_G} \cap x_{new}^{\tilde{*}_G} = Z. \end{aligned}$$

Proof. (1) Suppose any $x \in U$ can satisfy $x^{*D} = x_{new}^{\tilde{*}D}$. According to Theorem 2 and Definition 6, we have

$$\begin{aligned} x^{\tilde{*}_G \tilde{*}_G} &= \begin{cases} x^{*_G *_G} \cup \{x_{new}\}, & \text{if } x^{*_G} \subseteq x_{new}^{\tilde{*}_G}, \\ x^{*_G *_G}, & \text{otherwise,} \end{cases} \\ x^{\tilde{*}_D \tilde{*}_D} &= x^{*_D *_D} \cup \{x_{new}\}, \\ x^{*_G *_G} &\subseteq x^{*_D *_D}. \end{aligned}$$

Thus, $x^{\tilde{*}_G \tilde{*}_G} \subseteq x^{\tilde{*}_D \tilde{*}_D}$. Moreover, for $x_{new} \in U^+$, $x_{new}^{\tilde{*}_G \tilde{*}_G} = \{x_{new}\} \cup X \subseteq \{x_{new}\} \cup U = x_{new}^{\tilde{*}_D \tilde{*}_D}$, in which $X \subseteq U$ and for any $x \in X, x^{*_D} = x_{new}^{\tilde{*}_D}, x^{*_G} \cap x_{new}^{\tilde{*}_G} = x_{new}^{\tilde{*}_G}$.

Let $K \subset G$ be a granular consistent set of S^+. As G is the granular reduct of S, there must exist $x \in U$ such that $x^{*_K *_K} \nsubseteq x^{*_D *_D}$, which contradicts the assumption. Therefore, we conclude that G is a granular reduct of S^+.

(2) Suppose for any $x \in U$, $x^{*_D} \neq x_{new}^{\tilde{*}_D}$, $x^{*G} \cap x_{new}^{\tilde{*}_G} = Z$. According to Theorem 2 and Definition 6, we obtain

$$\begin{aligned} x^{\tilde{*}_G \tilde{*}_G} &= x^{*_G *_G}, \\ x^{\tilde{*}_D \tilde{*}_D} &= x^{*_D *_D}, \\ x^{*_G *_G} &\subseteq x^{*_D *_D}. \end{aligned}$$

Thus, $x^{\tilde{*}_G \tilde{*}_G} \subseteq x^{\tilde{*}_D \tilde{*}_D}$. As a result, for $x_{new} \in U^+$, $x_{new}^{\tilde{*}_G \tilde{*}_G} = \{x_{new}\} = x_{new}^{\tilde{*}_D \tilde{*}_D}$.

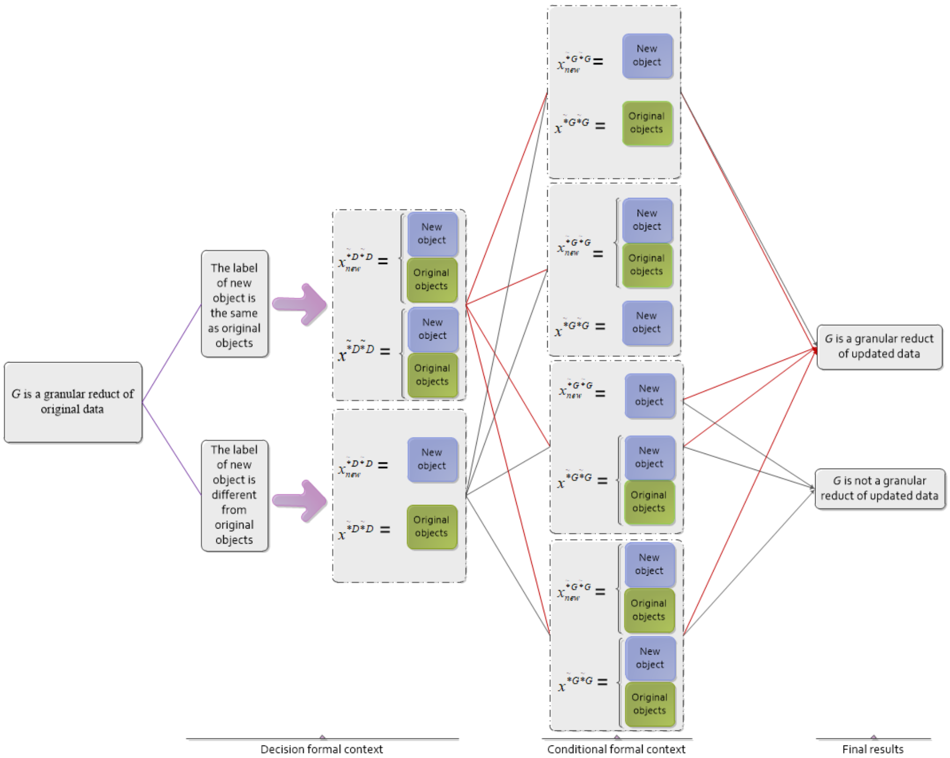

Figure 2: Consistency analysis of G after adding object x_{new}.

Assume that $K \subset G$ is a granular consistent set of S^+. As G is the granular reduct of S, there must exist $x \in U$ such that $x^{*_K *_K} \nsubseteq x^{*_D *_D}$, which is in contradiction with the assumption. Thus, we conclude that G is a granular reduct of S^+.

Theorem 4. Let $S = (U, C, I, D, J)$ and $S^+ = (U^+, C^+, I^+, D^+, J^+)$ be two formal decision contexts, G be a granular reduct of S and $E = C - G$. $G^+ = G \cup \{e\}, e \in E$ with $I(x, e) = 1$ and $I^+(x_{new}, e) = 0$ is a granular consistent set of S^+ if $x^{*_D} \neq x_{new}^{\tilde{*}_D}$ and $x^{*_G} \cap x_{new}^{\tilde{*}_G} = x^{*_G}$ for any $x \in U$.

Proof. Suppose any $x \in U$ can satisfy $x^{*_D} \neq x_{new}^{\tilde{*}_D}$ and $x^{*_G} \cap x_{new}^{\tilde{*}_G} = x^{*_G}$. According to Theorem 2, $x^{\tilde{*}_G\tilde{*}_G} = x^{*_G*_G} \cup \{x_{new}\}$ and $x^{\tilde{*}_D\tilde{*}_D} = x^{*_D*_D}$, which leads to the fact that G is not a granular consistent set of S^+. For any $e \in E$, $I(x, e) = 1$, $I^+(x_{new}, e) = 0$, $x^{*_{G^+}} \cap x_{new}^{\tilde{*}_{G^+}} = Z$. Then, according to Theorems 1-2, we get $x^{\tilde{*}_{G^+}\tilde{*}_{G^+}} = x^{*_{G^+}*_{G^+}} \subseteq x^{*_D*_D} = x^{\tilde{*}_D\tilde{*}_D}$. Consequently, $x_{new}^{\tilde{*}_{G^+}\tilde{*}_{G^+}} = \{x_{new}\} \subseteq x_{new}^{\tilde{*}_D\tilde{*}_D} = \{x_{new}\}$. Thus, G^+ is a granular consistent set of S^+.

Theorem 5. Let $S = (U, C, I, D, J)$ and $S^+ = (U^+, C^+, I^+, D^+, J^+)$ be two formal decision contexts, G be a granular reduct of S and $E = C - G$. $G^+ = G \cup \{e\}, e \in E$ with $I(x, e) = 0$ and $I^+(x_{new}, e) = 1$ is a granular consistent set of S^+ if $x^{*_D} \neq x_{new}^{\tilde{*}_D}$ and $x^{*_G} \cap x_{new}^{\tilde{*}_G} = x_{new}^{\tilde{*}_G}$ for any $x \in U$.

Proof. Suppose for any $x \in U$, $x^{*_D} \neq x_{new}^{\tilde{*}_D}$ and $x^{*_G} \cap x_{new}^{\tilde{*}_G} = x_{new}^{\tilde{*}_G}$. Thus, $x_{new}^{\tilde{*}_G\tilde{*}_G} = x_{new}^{*_G*_G} \cup \{x_{new}\}$ and $x_{new}^{\tilde{*}_D\tilde{*}_D} = \{x_{new}\}$, which yields that G is not a granular consistent set of S^+. For any $e \in E$, $I(x, e) = 0$, $I^+(x_{new}, e) = 1$, $x^{*_{G^+}} \cap x_{new}^{\tilde{*}_{G^+}} = Z$. Then, according to Theorems 1-2, we get $x^{\tilde{*}_{G^+}\tilde{*}_{G^+}} = x^{*_{G^+}*_{G^+}} \subseteq x^{*_D*_D} = x^{\tilde{*}_D\tilde{*}_D}$. Note that, $x_{new}^{\tilde{*}_{G^+}\tilde{*}_{G^+}} = \{x_{new}\} \subseteq x_{new}^{\tilde{*}_D\tilde{*}_D} = \{x_{new}\}$. Therefore, G^+ is a granular consistent set of S^+.

Theorem 6. Let $S = (U, C, I, D, J)$ and $S^+ = (U^+, C^+, I^+, D^+, J^+)$ be two formal decision contexts, G be a granular reduct of S and $E = C - G$. $G^+ = G \cup \{e, f\}, e, f \in E$ with $I(x, e) = 0, I^+(x_{new}, e) = 1$, $I(x, f) = 1$ and $I^+(x_{new}, f) = 0$ is a granular consistent set of S^+ if $x^{*_D} \neq x_{new}^{\tilde{*}_D}$ and $x^{*_G} = x_{new}^{\tilde{*}_G}$ for any $x \in U$.

Proof. Suppose for any $x \in U$, $x^{*_D} \neq x_{new}^{\tilde{*}_D}$ and $x^{*_G} = x_{new}^{\tilde{*}_G}$. Thus, $x^{\tilde{*}_G\tilde{*}_G} = x^{*_G*_G} \cup \{x_{new}\}$ and $x_{new}^{\tilde{*}_D\tilde{*}_D} = \{x_{new}\}$, which indicates that G is not a granular consistent set of S^+. For any $e, f \in E$, $I(x, e) = 0, I^+(x_{new}, e) = 1$, $I(x, f) = 1, I^+(x_{new}, f) = 0$, $x^{*_{G^+}} \cap x_{new}^{\tilde{*}_{G^+}} = Z$. Then, according to Theorems 1-2, we get $x^{\tilde{*}_{G^+}\tilde{*}_{G^+}} = x^{*_{G^+}*_{G^+}} \subseteq x^{*_D*_D} = x^{\tilde{*}_D\tilde{*}_D}$. Besides, $x_{new}^{\tilde{*}_{G^+}\tilde{*}_{G^+}} = \{x_{new}\} \subseteq x_{new}^{\tilde{*}_D\tilde{*}_D} = \{x_{new}\}$. Thus, G^+ is a granular consistent set of S^+.

Algorithm 1 The updating algorithm of a granular reduct

Input: A formal decision context S and newly added objects $U^{'}$.

Output: A granular consistent set G^{+} of the updated formal decision context S^{+}.

Initialize the granular reduct G for S

For each x_{new} in $U^{'}$

Set $K_a = \emptyset$; // The added attribute set

For each x_i in U

If x_i satisfies the conditions in Theorem 4 or Theorem 6

Select the candidate attribute set $K_s \subseteq C - G$ satisfying $I(x, e) = 1$ and $I^{+}(x_{new}, e) = 0$ for any $e \in K_s$;

Elseif x_i satisfies the conditions in Theorem 5 or Theorem 6

Select the candidate attribute set $K_v \subseteq C - G$ satisfying $I(x, b) = 0$ and $I^{+}(x_{new}, b) = 1$ for any $b \in K_v$;

End

If $K_a \cap K_s = \emptyset$

Select an attribute e from K_s;

Update $K_a = K_a \cup \{e\}$;

End

If $K_a \cap K_v = \emptyset$

Select an attribute b from K_v;

Update $K_a = K_a \cup \{b\}$;

End

End

End

$G^{+} = G \cup K_a$;

For each $e \in G^{+}$

If $G^{+} - \{e\}$ is a granular consistent set

$G^{+} = G^{+} - \{e\}$;

End

End

Return G^{+}.

Theorems 3-6 show the updating rules for the granular consistent sets in all possible cases when a new object is added. Roughly speaking, the changes in the granular reduct depend on the relationships between the conditional and decision attributes that the newly added objects and the previous objects have. In fact, one or two additional attributes are needed to maintain the consistency of an object in a new consistent formal decision context. We can design the updating algorithm (Algorithm 1) to compute the granular reduct based on those theorems. Note that, each theorem considers only a single case, such as Theorem 4 assuming that $x^{*_D} \neq x_{new}^{\tilde{*}_D}, x^{*_G} \cap x_{new}^{\tilde{*}_G} = x^{*_G}$ for any $x \in U$. Since knowledge updating is actually a complex process, two or more cases can exist at the same time. In other words, in addition to $x^{*_D} \neq x_{new}^{\tilde{*}_D}, x^{*_G} \cap x_{new}^{\tilde{*}_G} = x^{*_G}$, there can be another $y \in U$ that satisfies $y^{*_D} \neq x_{new}^{\tilde{*}_D}, y^{*_G} \cap x_{new}^{\tilde{*}_G} = x_{new}^{\tilde{*}_G}$, which leads to the coexistence of the cases introduced in Theorems 4-6. Therefore, it is necessary to prove that the updated granular consistent set can be suitable for all consistent decision formal contexts.

Theorem 7. Let $S = (U, C, I, D, J)$ and $S^+ = (U^+, C^+, I^+, D^+, J^+)$ be two formal decision contexts, G be a granular reduct of S and $E = C - G$. If we denote $K \subseteq E$ as the set of all the newly added attributes according to Theorems 4-6, then $G^+ = G \cup K$ is a granular consistent set of S^+.

Proof. Let $U_i \subseteq U, i = 1, 2, 3, 4$ be the object sets satisfying the conditions in Theorems 3-6, respectively, and $K_i \subseteq K$ be the corresponding attribute sets added by them.

Based on Proposition 1 and Definition 6, $x^{\tilde{*}_{G \cup K_i} \tilde{*}_{G \cup K_i}} \subseteq x^{\tilde{*}_G \tilde{*}_G} \subseteq x^{\tilde{*}_D \tilde{*}_D}$. Note that $G \cup K_i \subseteq G \cup K$. Thus, $x^{\tilde{*}_{G \cup K} \tilde{*}_{G \cup K}} \subseteq x^{\tilde{*}_{G \cup K_i} \tilde{*}_{G \cup K_i}}$. Then, for any $x \in U$, we have $x^{\tilde{*}_{G \cup K} \tilde{*}_{G \cup K}} \subseteq x^{\tilde{*}_D \tilde{*}_D}$.

Besides, for $x_{new} \in U^+$, $x_{new}^{\tilde{*}_{G \cup K} \tilde{*}_{G \cup K}} = \{x_{new}\} \cup X \subseteq \{x_{new}\} \cup U_1 = x_{new}^{\tilde{*}_D \tilde{*}_D}$, in which $X \subseteq U_1$ and for any $x \in X, x^{*_D} = x_{new}^{\tilde{*}_D}, x^{*_{G \cup K}} \cap x_{new}^{\tilde{*}_{G \cup K}} = x_{new}^{\tilde{*}_{G \cup K}}$.

In summary, for any $x \in U^+$, $x^{\tilde{*}_{G \cup K} \tilde{*}_{G \cup K}} \subseteq x^{\tilde{*}_D \tilde{*}_D}$, and G^+ is a granular consistent set of S^+.

3.2. Granular rule updating process

In this subsection, we investigate the updating of the granular rules extracted from a reduced formal decision context. Compared to the previous process, it is relatively simple. To obtain the updated granular rules, we only need to remove the attributes not included in the new granular reduct and add those included in the new granular reduct. Unless there are special notes, all of the granular rules to be discussed are extracted from the reduced formal decision context.

Theorem 8. Let $S = (U, C, I, D, J)$ and $S^+ = (U^+, C^+, I^+, D^+, J^+)$ be two formal decision contexts, and R_S, R_{S^+} be their granular rule sets, respectively. If they have the same granular reduct, then $R_{S^+} = R_S \cup \{r_{x_{new}}\}$.

Proof. It is obvious from Definition 4.

Theorem 9. Let $S = (U, C, I, D, J)$ and $S^+ = (U^+, C^+, I^+, D^+, J^+)$ be two formal decision contexts, G and G^+ be their granular reducts, and $R_S = \{x \in U | x^{*_G} \to x^{*_D}\}$ be the granular rule set of S. If R_S is updated to $R_{SU} = \{x \in U | x^{\tilde{*}_{G^+}} \to x^{\tilde{*}_{D^+}}\}$, then $R_{S^+} = R_{SU} \cup \{r_{x_{new}}\}$.

Proof. It is obvious from Theorem 8.

Algorithm 2 The updating algorithm of granular rules

Input: A formal decision context S and newly added objects U'.
Output: The granular rule set R_{S^+} of the updated formal decision context S^+.

Initialize the granular rule set R_S for S
For each x_{new} in U'
    Compute the granular reduct G^+ according to Algorithm 1;
    For each r_x in R_S
        Update $r_x : x^{*_G} \to x^{*_D}$ to $r_x : x^{\tilde{*}_{G^+}} \to x^{\tilde{*}_D}$;
        $R_{S^+} = R_{S^+} \cup \{r_x\}$;
    End
    If $x_{new}^{\tilde{*}_{G^+}} \to x_{new}^{\tilde{*}_D} \notin R_{S^+}$
        $R_{S^+} = R_{S^+} \cup \{x_{new}^{\tilde{*}_{G^+}} \to x_{new}^{\tilde{*}_D}\}$;
    End
End
Return R_{S^+}.

As Theorems 8-9 depicted, if the granular reduct does not change, the updating of the granular rules involves only the addition of a new rule. Otherwise, we need to adjust x^{*G} in r_x to $x^{\tilde{*}G^+}$ to generate a new granular rule. We summarize the updating progress in Algorithm 2, and give an example to illustrate the updating processes of a granular reduct and granular rules.

Example 1. We take Table 1 as an example to illustrate the whole updating process of a granular reduct and granular rules. Suppose that $U = \{1, 2\}$ is the initial object set and $U^+ = \{3, 4, 5, 6, 7, 8, 9, 10\}$ is the incremental object set. Then, the incremental processes are as follows:

Initialize the granular reduct $G = \{t, u\}$ and the granular rules:

$$r_{x_1} : \{t\} \to \{v\},$$
$$r_{x_2} : \{u\} \to \{v\}.$$

(1) For the new object x_3, we have $x_i^{*D} = x_3^{\tilde{*}D}(i = 1, 2)$. According to Theorem 3, $G_1^+ = G$ is a granular reduct of the updated formal decision context.

(2) For the new object x_4, we have $x_1^{*D} \neq x_4^{\tilde{*}D}, x_1^{*G} = x_4^{\tilde{*}G}$ and $x_i^{*D} \neq x_4^{\tilde{*}D}, x_i^{*G} \cap x_4^{\tilde{*}G} = Z(i = 2, 3)$. Based on Theorem 3 and Theorem 6, we need to select two attributes a_1, a_2 satisfying $I(x_1, a_1) = 1, I(x_4, a_1) = 0$, $I(x_1, a_2) = 0$ and $I(x_4, a_2) = 1$. Thus, $G_2^+ = G \cup \{c, d\} = \{c, d, t, u\}$ is a granular consistent set.

(3) For the new object x_5, we have $x_i^{*D} = x_5^{\tilde{*}D}(i = 1, 2, 3)$, $x_4^{*D} \neq x_5^{\tilde{*}D}$ and $x_4^{*G} \cap x_5^{\tilde{*}G} = Z$. According to Theorem 3, G_2^+ is a granular consistent set of the updated formal decision context.

(4) The granular reduct set G_2^+ is unchanged after the objects x_6 and x_7 are added into the updated formal decision context.

(5) For the new object x_8, we have $x_i^{*D} \neq x_8^{\tilde{*}D}, x_i^{*G} \cap x_8^{\tilde{*}G} = Z(i = 1, 2, 3, 5, 6)$, $x_j^{*D} \neq x_8^{\tilde{*}D}$ and $x_j^{*G} \cap x_8^{\tilde{*}G} = x_8^{\tilde{*}G}(j = 4, 7)$. According to Theorem 3 and Theorem 5, we need to select two attributes a_1, a_2 satisfying $I(x_4, a_1) = 0, I(x_8, a_1) = 1$, $I(x_7, a_2) = 0$ and $I(x_8, a_2) = 1$. Thus, $G_6^+ = G_2^+ \cup \{q\} = \{c, d, q, t, u\}$ is a granular consistent set.

(6) The granular reduct set G_6^+ is preserved after the objects x_9 and x_{10} are added into the updated formal decision context.

Finally, we can further obtain a granular reduct $G^+ = \{c, d, q, t\}$ from the granular consistent set $\{c, d, q, t, u\}$. Based on the updated granular reduct, the new granular rules are as follows:

$$r'_{x_1} : \{d, t\} \to \{v\},$$
$$r'_{x_2} : \{d\} \to \{v\},$$
$$r'_{x_4} : \{c, t\} \to \{w\},$$
$$r'_{x_8} : \{c, q\} \to \{x\}.$$

3.3. Prediction

In this section, we discuss the final prediction process. For an unlabelled object, the prediction stage aims at finding the most likely label for it. The key of this process is to choose the appropriate rule that matches the unlabelled object. Since the similarity degree can be used to measure the nearness between two objects, we employ a rule-based similarity degree to evaluate the closeness between two rules. Thus, some additional notions are needed.

Algorithm 3 Prediction algorithm for DRCM

Input: The trained rule base R_S and the testing formal decision context T.
Output: $l(T)$.

```
For each t_i in T
    For each r_x in R_s
        Calculate the similarity degree RS(r_{t_i}, r_x) according to Definition 9;
    End
    Get a label l(t_i) for the object t_i according to Definition 9 and Definition 10;
    l(T) = l(T) ∪ l(t_i);
End
Return l(T).
```

Definition 8. Let $S = (U, C, I, D, J)$ be a formal decision context, where $C = \{a_1, a_2, ..., a_m\}$. For any $x \in U$, we denote $v_x = (I(x, a_1), I(x, a_2), ..., I(x, a_m))$ as the attribute vector of x.

Definition 9. Let $S = (U, C, I, D, J)$ be a formal decision context. For two granular rules $r_x : x^{*_C} \rightarrow x^{*_D}$ and $r_y : y^{*_C} \rightarrow y^{*_D}$ of S, we define the granular rule-similarity degree between r_x and r_y as

$$RS(r_x, r_y) = \frac{v_x v_y^T}{|x^{*_C} \cup y^{*_C}|},$$

where v_y^T represents the transposition of v_y and $|X|$ is the number of elements in X. Since an attribute vector is 0 or 1 valued, $v_x v_y^T$ is actually equal to the number of the same elements of v_x and v_y.

With the definition of the granular rule-similarity degree, the issue of predicting labels can be transformed into the problem of determining the conclusion in a granular rule. For each unseen instance t in the testing set, there always exists a maximum RS denoted as $max(RS)$ and its corresponding granular rule $r_{\dot{x}}$. If we use $l(x)$ to denote the label of object x, then $l(t)$ is predicted by $l(\dot{x})$. However, sometimes a granular rule set marked by different labels can also be obtained. In this case, we cannot confirm which label is better. Thus, the following probability function is defined to offer a further prediction.

Definition 11. Let $S = (U, C, I, D, J)$ be a formal decision context, and R_S be the granular rule base trained by S. For an unseen instance t and a candidate set $R_1 = \{r_1, r_2, ..., r_n\} \subseteq R_S$ satisfying $RS(r_t, r_1) = RS(r_t, r_2) = ... = RS(r_t, r_n) = max(RS)$, we define

$$P(t, l(i)) = \frac{|\{r_k | r \in R_1, l(k) = l(i)\}|}{|R_1|} (i, k = 1, 2, ..., n)$$

as the probability of t labeled by $l(i)$.

Definition 11 indicates that the label with the highest probability actually corresponds to the label predicted by the trained granular rule base. In the prediction process, we first use the rule-similarity degree to predict a label for a given object; if it fails, then the probability must be used. If neither of them can predict a label effectively, it will be randomly selected from the candidate label set. The whole process is described by Algorithm 3.

3.4. Time complexity analysis

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

In Algorithm 1, Step 1 aims at initializing a granular reduct for S. Here, we use the algorithm proposed in this paper to initialize the granular reduct. Steps 2-19 select the candidate attribute set with a time complexity of $O(|U||U^{'}|)$, in which the judgements of $K_a \cap K_s = \emptyset$ and $K_a \cap K_v = \emptyset$ are to add as few attributes as possible. Steps 20-25 compute G^+ to evolve the granular consistent set for the purpose of generating a new granular reduct. Thus, the time complexity of Algorithm 1 is $O(|U||U^{'}| + |G^+|)$. In Algorithm 2, Step 1 initializes a granular rule set for S. Then, for each newly added object, it recalls Algorithm 1 to find a new granular reduct, which will severe as the basis of the updating granular rule in Steps 4-9. Since it also recalls Algorithm 1, its time complexity is $O(|U^{'}|(|U| + |G^+| + |R_S|))$. The time complexity of Algorithm 3 is $O(|T||R_S|)$.

4. Experimental results

In this section, we conduct experiments mainly for the purpose of evaluating the classification performances of the algorithms proposed in this paper.

4.1. Datasets

In the experiments, we chose sixteen datasets from the UCI Machine Learning Repository: Zoo DataSet, Iris Plants DataSet, Ecoli DataSet, Contraceptive Method Choice DataSet, Wisconsin Diagnostic Breast Cancer DataSet, Sensor Reading DataSet, Waveform Database Generator (Version 2) DataSet, Car Evaluation DataSet, Chess (King-Rook vs. King-Pawn)DataSet, Statlog (Image Segmentation) DataSet, Page Blocks Classification DataSet, Wilt DataSet, Mushroom DataSet, Spambase DataSet, Letter Recognition DataSet and Adult DataSet. Table 2 shows the statistics of the sixteen chosen datasets.

Table 2: The statistics of the chosen datasets

Dataset	#Object	#Attribute	#Class
Zoo DataSet	101	16	7
Iris DataSet	150	4	3
Ecoli DataSet	336	7	8
Contraceptive Method Choice DataSet	1473	9	3
Wisconsin Diagnostic Breast Cancer DataSet	569	30	2
Sensor Reading DataSet	5456	4	4
Waveform DataSet	5000	21	3
Car Evaluation DataSet	1728	6	4
Chess DataSet	3196	36	2
Statlog (Image Segmentation) DataSet	2310	19	7
Page Blocks Classification DataSet	5473	10	5
Wilt DataSet	4839	5	2
Spambase DataSet	4601	57	2
Mushroom DataSet	8124	21	2
Letter Recognition DataSet	20000	16	26
Adult DataSet	32561	14	2

Since most of the chosen datasets are many-valued or continuous-valued contexts, a data preprocessing technique should be applied to them to generate standard datasets. Conceptual scaling [48], which covers nominal scale, ordinal scale and interordinal scale, was employed to convert a many-valued or continuous-valued context into a 0-1 valued formal decision context. For example, the Wisconsin Diagnostic Breast Cancer DataSet has 30 continuous attributes. During the experiment, we first sorted the attribute values, and then divided them into three pairwise disjoint intervals to obtain an attribute with 3 attribute values only: intervals 1-3. Furthermore, we scaled each attribute into 3 attributes. Concretely, if the attribute value belongs to a certain interval, its value under the interval is 1; otherwise, it is 0. Thus, we obtained a standard dataset with 90 attributes. In addition, since this paper considers only consistent formal decision contexts, we eliminated the objects that cannot meet the consistency in the datasets, and renamed the standardized datasets by Datasets 1-16. Table 3 lists the details of the standardized datasets.

4.2. *Comparison of models*

To verify the effectiveness of DRCM, we compare it with two types of algorithms: the non-FCA-based algorithm and the FCA-based algorithm. Concretely, the former includes the Decision Tree (DT), Linear Discriminant Analysis (LDA), Linear Support Vector Machine (LSVM), Gaussian Kernel Function Support Vector Machine (GSVM), Polynomial Kernel Function Support Vector Machine (PSVM) and Radial Kernel Function Support Vector Machine (RSVM), Nearest Neighbour Classifiers [49] (KNN) with $k = 1$, Bagged Trees in Ensemble Classifiers [50] (EC), Random Forest (RF) and Naive Bayes (NB). In addition, considering that the latest CCLM [32] achieved the level of state-of-the-art in classification tasks, we also compare DRCM and CCLM. For fairness, all of the classifiers were trained automatically by the classification learner software tool in MATLAB 2016b, and 5 trials were conducted to report the average value with the same training and testing samples. Statistical analysis was performed by using paired t-test[1] on MATLAB to compare the two algorithms, and the test results are shown in the appropriate place. The experimental conditions were as follows: Intel Core i7-4720HQ @2.60 GHz CPU and 16 GB main memory.

For comparison, we will first analyse the feasibility of the proposed algorithm for achieving classification tasks, and then, we will compare it with the classification algorithms introduced above; next, we will evaluate the dynamic learning ability. Since the classification task aims at predicting the labels of new objects with known objects, it requires a training set and testing set. Thus, during the experiment, a given dataset will be randomly divided into a training set and testing set, and $TestRatio = \frac{|testing\ set|}{|training\ set|+|testing\ set|}$ is employed to depict the division ratio.

[1] https://uk.mathworks.com/help/stats/ttest.html#d123e816945

Table 3: The statistics of the datasets after standardization

Dataset	#Object	#Attribute	#Class
Dataset 1	93	16	7
Dataset 2	105	12	3
Dataset 3	199	17	8
Dataset 4	254	21	3
Dataset 5	569	90	2
Dataset 6	660	12	4
Dataset 7	1306	21	3
Dataset 8	1590	19	4
Dataset 9	1856	38	2
Dataset 10	2070	55	7
Dataset 11	2132	30	5
Dataset 12	4067	15	2
Dataset 13	4125	171	2
Dataset 14	8000	50	2
Dataset 15	8871	48	26
Dataset 16	15095	40	2

Table 4: The change of classification ability (%) after attribute reduction

Dataset	TestRatio				
	0.1	0.2	0.3	0.4	0.5
Dataset 1	+2.2222	+0.0000	-0.7407	-0.5440	-3.0434
Dataset 2	+0.0000	+0.9524	+0.6452	+0.0000	+0.3847
Dataset 3	+0.0000	+0.0000	+1.0170	-0.2531	+0.6061
Dataset 4	+0.8000	+1.6000	+2.3685	+1.9802	+0.6300
Dataset 5	-0.3571	-0.7079	-1.5294	-0.9691	-2.3943
Dataset 6	+0.5051	+0.3031	+0.5051	+0.0000	+0.3031
Dataset 7	-1.2307	+0.0000	-2.8644	-3.4099	-6.0030
Dataset 8	+0.0000	+0.0000	+0.0000	+0.0000	+0.0000
Dataset 9	+0.1082	+0.0001	-0.0719	-0.0269	-0.1939
Dataset 10	-0.3864	-0.2415	-0.7085	-0.1932	-0.6570
Dataset 11	+0.0000	+0.0939	+0.0627	-0.1643	+0.1877
Dataset 12	+0.0000	+0.1477	+0.0164	-0.0123	-0.1475
Dataset 13	+0.6739	+0.6848	+0.7899	+0.8348	+0.5826
Dataset 14	+0.0000	+0.0000	+0.0000	+0.0000	+0.0000
Dataset 15	+0.0000	+0.0000	+0.0000	+0.0000	+0.0000
Dataset 16	-0.0795	+0.0596	-0.0441	+0.0000	+0.0000

(i) Feasibility analysis of the proposed algorithm

Feasibility analysis evaluates whether the proposed attribute reduction method can retain the effective information of the datasets. The classification ability of the dataset is used as the evaluation standard in this paper. In other words, a good attribute reduction algorithm should be able to maintain the classification ability of the dataset. For this purpose, we calculated statistics on the changes in the classification ability under different partition strategies, which are reported in Table 4.

It can be observed that attribute reduction proposed in this paper guarantees the classification ability on Datasets

Table 5: Training time of DRCM

Dataset	#Training	#Testing	Time (s)	Ave. (obj.)	Ave. (att.)
Dataset 1	75	18	0.0235	0.0003	0.0011
Dataset 2	84	21	0.0074	0.0001	0.0006
Dataset 3	160	39	0.0425	0.0002	0.0025
Dataset 4	204	50	0.1264	0.0006	0.0060
Dataset 5	456	113	0.8034	0.0017	0.0089
Dataset 6	528	132	0.0131	0.0002	0.0010
Dataset 7	1045	261	2.4491	0.0023	0.1166
Dataset 8	1383	345	0.7872	0.0037	0.0414
Dataset 9	1485	371	7.0439	0.0005	0.1853
Dataset 10	1656	414	6.5984	0.0032	0.1199
Dataset 11	1706	426	0.4131	0.0039	0.0130
Dataset 12	3254	813	0.0819	0.0001	0.0054
Dataset 13	3300	825	119.4784	0.3620	0.6987
Dataset 14	6400	1600	6.5885	0.0010	0.1317
Dataset 15	7097	1774	49.1988	0.0069	1.0249
Dataset 16	12076	3019	113.5600	0.0094	2.8390

8, 14 and 15, while it improves the classification ability on Datasets 2, 4, 6 and 13. On most of the remaining datasets, the proposed attribute reduction method can also improve the classification ability of the obtained rules with an appropriate partition. For example, the classification ability on Dataset 3 has been improved by 1.0170% when TestRatio = 0.4, which indicates that attribute reduction is helpful for us to improve classification ability. This finding occurs because attribute reduction can reduce some attributes that have adverse effects on the classification task, i.e., an increase in the associated rule-similarity degree. Despite the effectiveness of DRCM, it performed poorly on some of the datasets, such as a 6.003% descent on Dataset 7, which could be caused by an inappropriate granular reduct. Here, we only want to show that an appropriate granular reduct can indeed improve the generalization ability of the proposed model.

The training time and the average value are also reported in Table 5. Obviously, the training time is closely related to the numbers of objects and attributes. Thus, the proposed algorithm can train at a very fast speed in the face of small datasets, and only 0.0235 s was needed to train a formal decision context with 75 objects and 18 attributes. When facing a much larger dataset with 12076 objects and 40 attributes, the proposed algorithm can also be quickly completed in 114 seconds, which shows the feasibility of the proposed algorithm.

(ii) A comparison with other classification algorithms

To achieve the comparison task, we must report the experimental results when compared with the non-FCA based algorithms. The prediction accuracies with TestRatio = 0.2 are listed in Table 6, from which we can see that DRCM achieves the optimal results in eight of the sixteen chosen datasets. In other datasets, it still obtains competitive results. To further analyse the influence of the partition strategy on the proposed algorithm, we draw accuracy trend charts of DRCM on the sixteen datasets with different TestRatios in Figure 3. It is obvious that the accuracy of DRCM fluctuates slightly with the increase in TestRatio. Since only incomplete granular rules can be learned from a small number of instances and noise can be produced by a large number of instances, selecting an appropriate partition ratio is very important for our algorithm.

Moreover, we show the prediction accuracy of DRCM and the simplified CCLM [32] in Figure 4. DRCM has higher accuracies on all of the datasets except for Datasets 1 and 5-7, and even improves the accuracy by up to 4% on Dataset 4. In addition, our model displays more stable results than CCLM in some datasets with increasing of TestRatio. A further theoretical analysis of the stable results is performed as follows. It appears that there is little diversity between some objects on these datasets, which leads to more candidate labels by calculating the similarity degree, and those candidate labels have a negative impact on the final results of CCLM, while DRCM reduces some of such attributes and increases the diversity between objects, which makes our model have better prediction results

Table 6: Accuracy (%) comparison with non-FCA based algorithms

Dataset	DRCM	DT	LAD	LSVM	GSVM	KNN	EC	RF	NB	PSVM	RSVM
Dataset 1	**97.7778**	92.2222	95.5556	94.4444	95.5556	95.5556	96.6667	96.6667	94.4444	96.6667	96.6667
Dataset 2	**100.000**	99.0476	95.2381	**100.000**	**100.000**	**100.000**	99.0476	**100.000**	98.0952	99.0476	**100.000**
Dataset 3	**94.3590**	93.3333	92.3077	93.3333	93.8462	91.7949*	93.3333	93.3333	91.7949	92.8205	93.8462
Dataset 4	72.4000	70.8000	64.0000*	70.4000	**73.6000**	70.0000*	70.8000*	70.0000	68.8000*	74.0000	**73.6000**
Dataset 5	94.5133	94.8673	95.7522	97.8761*	97.8761*	94.5133	96.9912*	97.5221*	94.6903	97.3451*	**97.8761***
Dataset 6	**98.4848**	98.3333	90.0000	97.5758*	98.0303	98.0303	98.3333	97.5757	97.1212*	98.1818	98.0303
Dataset 7	86.7433	88.8889*	93.7165*	93.4099*	93.7931	88.0460*	92.4904*	92.0306*	93.4100*	92.7203*	**93.8697***
Dataset 8	**99.8742**	97.6101*	91.8868*	95.9748*	99.8113	95.6604*	99.4340	99.4969	88.8050*	99.8113	99.8113
Dataset 9	98.0054	99.5148*	90.3504*	99.0296*	98.7601	95.7952*	99.7305*	**99.7844***	89.9191*	98.7601	98.7601
Dataset 10	94.2995	94.8309	84.3961*	93.9614	94.8309	93.3333	95.7971*	96.0869*	84.9275*	**96.4735***	94.8309
Dataset 11	**98.4037**	97.4178*	92.6291*	97.7934*	97.9812*	97.9813*	98.2629	98.1221	94.0376*	98.2629	98.0282*
Dataset 12	**98.3764**	98.2780	97.7614*	98.2780	98.2780	98.0074	98.2780	98.3026	96.8758	98.3518*	98.2780
Dataset 13	91.5394	91.1030	90.6182*	93.0909*	94.0364*	90.8848	94.2788*	**94.3273***	91.9758	94.1333*	94.1818*
Dataset 14	**100.000**	95.4000	96.0375*	74.4625*	72.0875*	74.7875*	**100.000**	95.6750*	**100.000**	**100.000**	71.8000*
Dataset 15	88.3653	64.4194*	63.3371*	81.9504	87.9030*	86.7531*	88.8162*	**89.1545***	66.9335	88.0045*	88.0158*
Dataset 16	94.8062	95.3097*	94.5545	95.3694	95.6807*	94.3557*	**95.9987***	95.8794*	94.6538	95.4952*	95.6608*

The best results are highlighted, and "*" indicates that DRCM is better (worse) than the corresponding algorithm according to the paired t-test with confident level of 95 percent.

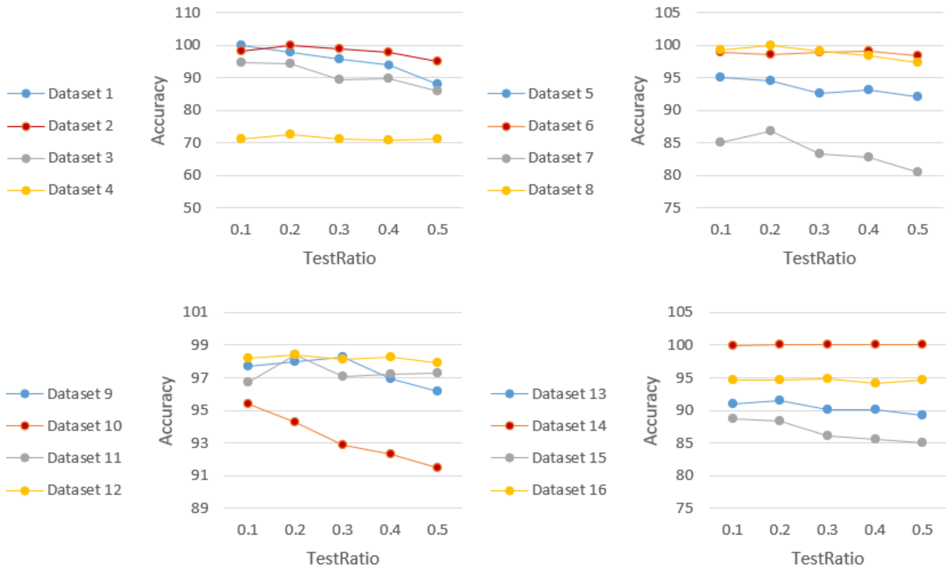

Figure 3: Evaluation results of DRCM with different partition strategies.

due to reducing the negative impact caused by unnecessary attributes. However, DRCM achieves poorer results in the five partition strategies on Datset 1. This finding could occur because of the strong correlations between attributes, which leads to the fact that rules have less information than concepts after reduction.

(iii) Performance evaluation of the dynamic learning ability

Dynamic learning aims to evaluate the ability of a model to address new data. Since DRCM is an updating-based model, it has a good theoretical basis for dynamic learning. In what follows, we will conduct a performance evaluation

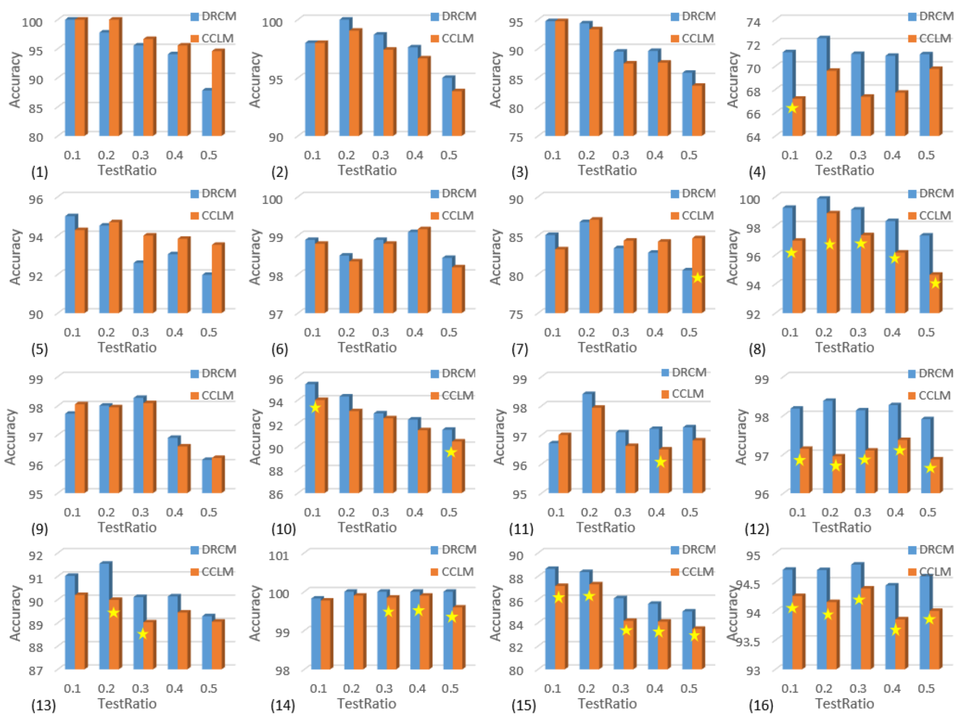

Figure 4: Evaluation results of DRCM and CCLM.
The five-pointed star means that DRCM is better (worse) than CCLM acccording to the paired t-test with confident level of 95 percent.

Table 7: The batch size of different batches

Dataset	Batch 1	Batch 2	Batch 3	Batch 4	Batch 5
Dataset 1	15	21	20	19	18
Dataset 2	16	23	23	22	21
Dataset 3	32	43	43	42	39
Dataset 4	40	55	55	54	50
Dataset 5	91	123	122	120	113
Dataset 6	105	142	142	139	132
Dataset 7	209	281	280	275	261
Dataset 8	254	342	341	335	318
Dataset 9	297	398	398	392	371
Dataset 10	331	444	444	437	414
Dataset 11	341	458	457	450	426
Dataset 12	650	873	872	859	813
Dataset 13	660	885	884	871	825
Dataset 14	1280	1716	1715	1689	1600
Dataset 15	1419	1903	1902	1873	1774
Dataset 16	2415	3237	3236	3188	3019

of dynamic learning ability on the chosen datasets. To generate a dynamic dataset, those datasets were divided into five batches named Batches 1-5, in which Batch 1 served as the initial dataset, Batches 2-4 were regarded as the incremental dataset, and Batch 5 was the testing dataset. Table 7 lists the batch size of the batches on each dataset.

Table 8 shows the evaluation results for dynamic learning, in which Batches 1,2-3-5 indicate that Batch 1 together

Table 8: The evaluation results for dynamic learning

Dataset	Batches 1-5		Batches 1-2-5		Batches 1,2-3-5		Batches 1,2,3-4-5	
	Time(s)	Accuracy(%)	Time(s)	Accuracy(%)	Time(s)	Accuracy(%)	Time(s)	Accuracy(%)
Dataset 1	0.0073	76.6667	0.0088	90.0000*	0.0136	93.3333*	0.0161	**96.6666***
Dataset 2	0.0063	80.9520	0.0058	94.2857*	0.0055	**98.0952***	0.0062	98.0952*
Dataset 3	0.0108	76.4102	0.0171	84.6154	0.0231	88.2051*	0.0302	**89.2308***
Dataset 4↓	0.0234	67.2000	0.0429	71.2000	0.0791	**74.8000***	0.1239	71.6000
Dataset 5↓	0.0492	88.1416	0.1888	89.9115	0.4380	**92.7434***	0.7258	92.2124*
Dataset 6	0.0091	97.7273	0.0104	97.8789	0.0116	**97.8788**	0.0117	**97.8788**
Dataset 7↓	0.2078	67.1264	0.9877	77.0115*	2.4901	**82.2988***	4.3077	79.9234*
Dataset 8	0.1387	90.8176	0.3631	96.3522*	0.6031	99.3082*	0.7719	**99.6226***
Dataset 9	0.4121	92.7763	1.8070	95.4178*	4.1424	97.0889*	7.0647	**97.7897***
Dataset 10	0.4740	87.1015	1.9048	91.4010*	3.8877	93.2367*	6.1497	**94.4444***
Dataset 11	0.0965	95.6338	0.2117	96.7136	0.3178	97.3709*	0.4282	**97.5117***
Dataset 12↓	0.0389	98.3518	0.0499	98.4010	0.0714	**98.4256**	0.0755	98.4010
Dataset 13	5.0832	84.1697	28.4248	88.4849*	66.8538	90.2545*	126.8750	**91.3212***
Dataset 14↓	1.5775	**98.9162**	3.7597	98.8177	5.3921	98.5344	6.6409	98.5714
Dataset 15	5.6487	78.0045	19.0140	84.1826*	36.9245	86.6178*	56.4153	**88.2863***
Dataset 16	9.2369	93.9715	40.0850	94.2762*	79.4134	94.5478*	125.9319	**94.7333***

The best results are highlighted, and "*" indicates that the result is better (worse) than the initial state according to the paired t-test with confident level of 95 percent.

with Batch 2 serves as the initial dataset while Batch 3 and Batch 4 are regarded as the incremental data and testing set, respectively. During the incremental process, the datasets that suffer from accuracy loss are labelled by "↓". Obviously, the accuracies of DRCM are nondecreasing on all of the datasets except for Datasets 4, 5, 7, 12 and 14 as the increment process continues. In addition, DRCM achieved the highest accuracy on Dataset 14 in the first process, obtained the best results on Datasets 2, 4-7 and 12 in the second process, and obtained the best results on the remaining datasets in the third process. In other words, DRCM indeed has better incremental learning ability.

We also evaluated the impact of batch size on the dynamic learning process, and the results are shown in Figure 5. The prediction accuracy can be slightly improved with increasing batch size on most of the datasets. However, at the same time, it should be pointed out that a batch that is too large could also lead to a decline in the prediction ability. In fact, this circumstance is exactly consistent with the gradual cognitive process. On the one hand, a larger batch size will bring challenges to cognitive ability, and on the other hand, it could produce more noise. In other words, how to select a proper batch size is crucial to our model.

5. Conclusions and future work

In this paper, we have studied dynamic knowledge discovery in a formal decision context. Concretely, a novel model named DRCM has been proposed for updating granular reduct and granular rules by analysing the newly added objects. We performed a feasibility analysis of the proposed method and compared it with other popular classification algorithms, as well as the evaluation of the dynamic learning ability. The experimental results have illustrated that the proposed DRCM can achieve better classification performance on many datasets.

Although DRCM has shown good experimental performance, it still has some limitations. First, the training time increases with increasing data volume, and thus, it could take a long time to train a large-scale dataset. Therefore, optimization methods, such as using parallel technology, for the updating process are still worthwhile to explore. Second, DRCM was discussed for 0-1 valued contexts, which cannot be able to address continuous data directly. The discretization of continuous attributes often means information loss. Thus, DRCM still deserves to be investigated in a fuzzy environment to improve the scalability of the model. Moreover, the consistent condition for the formal

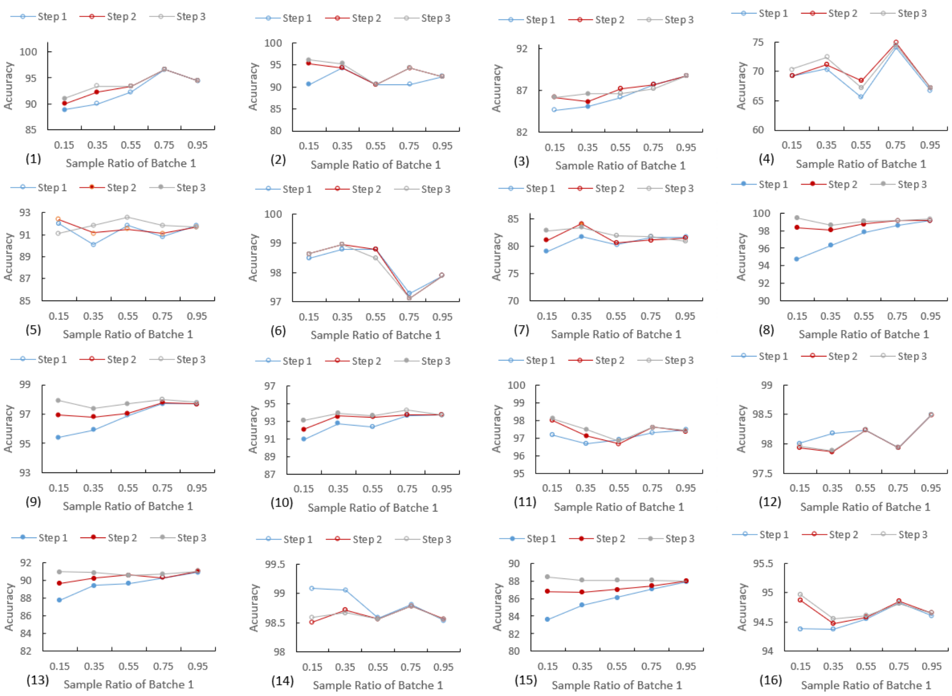

Figure 5: Evaluation results of dynamic learning with different sample ratios
The "•" indicates that result of the incremental process is better (worse) than the initial state according to the paired t-test with confident level of 95 percent.

decision context makes DRCM lack effectiveness in the face of inconsistent data. In other words, how to avoid the inefficiency caused by inconsistency is also worthwhile to discuss. Since each dataset must contain consistent subsets, this problem can be solved by finding a consistent subset to obtain a local granular reduct. This concern is another focus in our future work. Last but not least, due to the non-exclusiveness of granular reduct, the performance of our model after reduction could become slightly unsatisfactory, and thus, developing an improved DRCM that can overcome this limitation is also an important area to be discussed.

Credit authorship contribution statement

Jiaojiao Niu: Software, Data curation, Writing-Original draft preparation, Visualization. **Degang Chen**: Writing-Reviewing and Editing, Funding acquisition. **Jinhai Li**: Investigation, Writing-Reviewing and Editing. **Hui Wang**: Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Ishibuchi, K. Nozaki, H. Tanaka, Distributed representation of fuzzy rules and its application to pattern classification, Fuzzy Sets and Systems 52 (1) (1992) 21-32.
[2] H. Ishibuchi, K. Nozaki, H. Tanaka, Efficient fuzzy partition of pattern space for classification problems, Fuzzy Sets and Systems 59 (3) (1993) 295-304.

[3] K. Nozaki, H. Ishibuchi, H. Tanaka, Adaptive fuzzy rule-based classification systems, IEEE Transactions on Fuzzy Systems 4 (3) (1996) 238-250.

[4] C. T. Sun, Rule-base structure identification in an adaptive-network-based fuzzy inference system, IEEE Transactions on Fuzzy Systems 2 (1) (1994) 64-73.

[5] R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, in: Ordered Sets, I. Rival, Ed. Dordrecht, The Netherlands: Springer, 1982, pp. 445-470.

[6] M. E. Cintra, M. C. Monard, H. A. Camargo, FCA-based rule generator, a framework for the genetic generation of fuzzy classification systems using formal concept analysis, in: IEEE International Conference on Fuzzy Systems, 2015, pp. 1-8.

[7] K. S. Tiwari, A. G. Kothari, Attribute reduction algorithm for inconsistent information system using rough set theory, in: Third International Conference on Computational Intelligence and Information Technology (CIIT 2013), 2013, pp. 218-224.

[8] D. Ślęzak, Approximate reducts in decision tables, in: Proceedings of the Sixth International Conference, Information Procesing and Management of Uncertainty in Knowledge Based Systems (IPMU'96), 1996, pp. 1159-1164.

[9] Y. P. Qu, Z. Xu, C. J. Shang, et al., Inconsistency guided robust attribute reduction, Information Sciences 580 (2021) 69-91.

[10] Y. H. She, Z. H. Qian, X. L. He, et al., On generalization reducts in multi-scale decision tables, Information Sciences 555 (2021) 104-124.

[11] K. Y. Liu, X. B. Yang, H. Fujita, et al., An efficient selector for multi-granularity attribute reduction, Information Sciences 505 (2019) 457-472.

[12] Z. Pawlak, Rough sets, International Journal of Computer and Information Sciences 11 (1982) 341-356.

[13] Z. Pawlak, A. Skowron, Rough sets: some extensions, Information Sciences 177 (1) (2007) 28-40.

[14] L. A. Zadeh, Fuzzy sets, Information & Control 8 (3) (1965) 338-353.

[15] J. G. Bazan, Dynamic reducts and statistical inference, in: Proceedings of the Sixth International Conference, Information Procesing and Management of Uncertainty in Knowledge Based Systems (IPMU'96), 1996, pp. 1147-1152.

[16] J. G. Bazan, A. Skowron, Dynamic reducts as a tool for extracting laws from decisions tables, in: Proceedings of the Eighth International Symposium on Methodologies for Intelligent Systems (ISMIS'94), Lecture Notes in Artificial Intelligence 869. Berlin: Springer-Verlag 1994, pp. 346-355.

[17] J. G. Bazan, A. Skowron, D. Ślęzak, et al., Searching for the complex decision reducts: the case study of the survival analysis, in: N. Zhong, Z. Ras, S.Tsumoto, E. Suzuki (Eds.), Proceedings of the 14th International Symposium on Methodologies for Intelligent Systems, 2871, Springer, 2003, pp. 160-168.

[18] F. M. Brown, Boolean reasoning: the logic of Boolean equations, Kluwer, Dordrecht, 1990.

[19] A. Skowron, C. Rauszer, The discernibility matrices and functions in information systems, in: R. Slowinski (Ed.), Intelligent Decision Support: Handbook of Applications and Advances of Rough Sets Theory, Kluwer Academic Publisher, Dordrecht, 1992, pp. 331-362.

[20] W. X. Zhang, G. F. Qiu, Uncertain decision making based on rough sets, Tsinghua University Press, 2005.

[21] L. Wei, J. J. Qi, W. X. Zhang, Attribute reduction theory of concept lattice based on decision formal contexts, Science in China Series F: Information Sciences 51 (7) (2008) 910-923.

[22] W. Z. Wu, Y. Leung, J. S. Mi, Granular computing and knowledge reduction in formal contexts, IEEE Transactions on Knowledge and Data Engineering 21 (10) (2009) 1461-1474.

[23] M. Liu, M. W. Shao, W. X. Zhang, et al., Reduction method for concept lattices based on rough set theory and its application, Computers & Mathematics with Applications 53 (9) (2007) 1390-1410.

[24] Y. Y. Yao, Three-way decisions and cognitive computing, Cognitive Computation 8 (2016) 543-554.

[25] Y. Y. Yao, Three-way decision and granular computing, International Journal of Approximate Reasoning 103 (2018) 107-123.

[26] W. Pedrycz, Granular computing: analysis and design of intelligent systems, CRC Press, 2013.

[27] W. Pedrycz, Granular computing for data analytics: a manifesto of human-centric computing, IEEE/CAA Journal of Automatica Sinica 5 (6) (2018) 1025-1034.

[28] L.Wei, Q.Wan, Granular transformation and irreducible element judgment theory based on pictorial diagrams, IEEE Transactions on Cybernetics 46 (2) (2016) 380-387.

[29] J. H. Li, C. C. Huang, J. J. Qi, et al., Three-way cognitive concept learning via multi-granularity, Information Sciences 378 (2017) 244-263.

[30] J. M. Ma, M. J. Cai, C. J. Zou, Concept acquisition approach of object-oriented concept lattices, International Journal of Machine Learning and Cybernetics 8 (1) (2017) 123-134.

[31] J. H. Li, C. L. Mei, W. H. Xu, et al., Concept learning via granular computing: a cognitive viewpoint, Information Sciences 298 (1) (2015) 447-467.

[32] Y. Shi, Y. L. Mi, J. H. Li, et al., Concept-cognitive learning model for incremental concept learning, IEEE Transactions on Systems, Man, and Cybernetics: Systems 52 (2021) 809-821.

[33] S. O. Kuznetsov, Complexity of learning in concept lattices from positive and negative examples, Discrete Applied Mathematics 142 (1-3) (2004) 111-125.

[34] 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 9 (2018) 1785-1805.

[35] H. M. Chen, T. R. Li, D. Ruan, et al., A rough-set-based incremental approach for updating approximations under dynamic maintenance environments, IEEE Transactions on Knowledge and Data Engineering 25 (2) (2013) 274-284.

[36] Y. G. Jing, T. R. Li, H. Fujita, et al., An incremental attribute reduction method for dynamic data mining, Information Sciences 465 (2018) 202-218.

[37] Y. Y. Huang, T. R. Li, C. Luo, et al., Matrix-based dynamic updating rough fuzzy approximations for data mining, Knowledge-Based Systems 119 (C) (2017) 273-283.

[38] W. H. Xu, J. Z. Pang, S. Q. Luo, A novel cognitive system model and approach to transformation of information granules, International Journal of Approximate Reasoning 55 (3) (2014) 853-866.

[39] J. G. Bazan, A. Skowron, R. W. Swiniarski, Rough sets and vague concept approximation: from sample approximation to adaptive learning, in: J. F. Peters, A. Skowron (Ed.), Transactions on Rough Sets V. Lecture Notes in Computer Science, 4100. Springer, Berlin, Heidelberg,

2006, pp. 39-62.
[40] T. Y. Lin, Y. Y. Yao, L. A. Zadeh, Data mining, rough sets and granular computing, Heildelberg, Germany: Physica-Verlag, 2002.
[41] B. B. Sang, H. M. Chen, T. R. Li, et al., Incremental approaches for heterogeneous feature selection in dynamic ordered data, Information Sciences 541 (2020) 475-501.
[42] P. Ni, S. Y. Zhao, X. Z. Wang, et al., Incremental feature selection based on fuzzy rough sets, Information Sciences 536 (2020) 185-204.
[43] W. H. Xu, W. T. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets, IEEE Transactions Cybernetics 46 (2) (2016) 366-379.
[44] L. A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 90 (2) (1997) 111-127.
[45] Y. Q. Zhang, Constructive granular systems with universal approximation and fast knowledge discovery, IEEE Transactions on Fuzzy Systems 13 (1) (2005) 48-57.
[46] A. Skowron, A. Jankowski, Interactive granular computing, in: D. Ciucci, G. Wang, S. Mitra, W. Z. Wu (Ed.), Rough Sets and Knowledge Technology. RSKT 2015. Lecture Notes in Computer Science, 9436. Springer, Cham 2015, pp. 50-61.
[47] A. Skowron, M. Szczuka, Toward interactive computations: a rough-granular approach, Advances in Machine Learning II (2010) 23-42.
[48] B. Ganter, R. Wille, Formal concept analysis: mathematical foundations, Springer-Verlag, New York, Inc, 1999.
[49] T. M. Cover, P. E. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13 (1) (1967) 21-27.
[50] L. Rokach, Ensemble-based classifiers, Artificial Intelligence Review 33 (2010) 1-39.