



# A new mechanism of rule acquisition based on covering rough sets

Xiaoping Zhang<sup>1</sup> · Jinjin Li<sup>2</sup> · Weikang Li<sup>3</sup>

Accepted: 30 November 2021 / Published online: 4 February 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Rule acquisition, known as knowledge acquisition, is an important and topical issue in granular computing theory. Granules are not only composed of objects but also have feature values. However, In the granule associativity rules, the traditional rule extraction methods fail to consider the influence of granules on the decision, thus the method is not well adapted in reality. On the other hand, the existing methods lack a rule-based measure for information systems. In this paper, the action parameters are first introduced to establish a more realistic granule associativity rule in covering information system. Further, we present the rule-based data potential to address the measurement problem. In addition, rule-based scale selection in multi-scale covering rough sets is explored, followed by a scale combination integrating generalization capability, data potential, and lower approximation. Finally, algorithm is designed and experiments are conducted.

**Keywords** Covering information system · Multi-scale · Rule acquisition · Association rule

## 1 Introduction

Granular computing, a data processing method, has become a popular research direction [17]. In granular computing, granule is the basic unit for solving problems and generally known as an indivisible element or subset of the set [23]. Thus, the main research problem of granular computing is how to construct and use information granules to calculate [27]. Different level and thickness of granules have different influence on the problem solving process. For example, we will gain smaller knowledge granularity if the description of an object is more careful. However, we will gain the lower generalization ability in acquiring knowledge and greater complexity in solving the problem at the same time.

Information system is a common carrier used for granular computing [19]. In order to deal with incomplete and fuzzy data, Pawlak proposed the classical rough set theory based

on the idea that upper and lower approximation operators can approach the target set [12, 13]. In data mining, we do not need any priori knowledge using rough set theory. Thus, the theory has attracted considerable attention in the area of computer and mathematics since it was proposed [16, 30]. At present, the theory and its innovative extensions have been widely used in data mining, medical diagnosis, decision analysis, feature selection and so on [14].

Classical rough set is based on equivalence relation [12, 13]. But an object may belong to multiple classes in practice. Therefore, Zakowski proposed the covering rough set in 1983 [24]. Covering rough set is one of the important generalized models. Many scholars constructed different upper and lower approximation operators of covering rough set in different perspectives [1, 4, 14, 15, 20–22, 29–32]. However, these studies are limited to single-scale covering. Subsequently, some scholars extended the covering rough set from single scale to multi-scale [7–10, 25]. It is more comprehensive and flexible to apply multi-scale covering rough set to data mining. In multi-scale covering rough set, not only knowledge acquisition at different levels but also knowledge and granularity conversion between different levels can be realized.

In data mining, rule extraction serves as knowledge discovery [19, 27]. A rule contains predecessor and successor, i.e. the condition and decision. The certain factor of the rule is essentially the degree to which condition granule supports decision granule. According to the idea of rough set theory,

✉ Jinjin Li  
jinjinli@mnnu.edu.cn

<sup>1</sup> School of Mathematics Sciences, Huaqiao University, 362021, Fujian, Quanzhou, China

<sup>2</sup> Fujian Province University Key Laboratory of Computational Science, School of Mathematics Sciences, Huaqiao University, 362021, Fujian, Quanzhou, China

<sup>3</sup> School of Economics and Finance, Huaqiao University, 362021, Fujian, Quanzhou, China

the lower approximation represents certain classification and certain rules, while the upper approximation represents uncertain information and uncertain (i.e., possible) rules [18]. There have been many studies on rule extraction [2, 5, 6, 11, 18, 26, 28].

In this paper, we firstly propose a rule validity function to reflect the classical rule and promote it to fit the granule associativity rule. Then motivated by the adaptability of rule extraction methods and the measurement of decision descriptions, more effective granule associativity rules are established by considering the impact of different granules on decision descriptions. There are three types of objects holding intensities of rules, i.e., strongly, generally and barely. Further, the rule-based data potential of information systems for decision descriptions is reflected by considering the rule-based classification of objects and then measuring the number of objects. We mainly use it to evaluate the potential of information systems. Subsequently, we explore the use of the new rule extraction mechanism in scale selection. The optimal scale combination based on rule extraction is put forward for the first time. The final work is to conduct experiments on three UCI datasets [3], including Iris Plants, Hayes-Roth and Somerville Happiness Survey. As a result, we present a new mechanism of rule acquisition based on covering rough sets, which can be more effectively fitted to the reality than the classical rule extraction.

The rest parts of the paper are organized as follows. Some related work on our proposed mechanism of covering rough set and rule acquisition is introduced in Section 2. In Section 3, some basic notions of covering rough set and rules extraction are reviewed, and some notations are explained. In Section 4, the mechanism of rule extraction is studied, and rule-based data potential is proposed. In Section 5, the proposed mechanism of rule acquisition is employed for scale selection. In Section 6, algorithm for calculation of rule-based data potential is given, and several experiments are shown. Finally, we wrap up the paper with a summary and an outlook of further research in Section 7.

## 2 Related work

In this section, we mainly review some related studies in covering rough set and rule extraction.

Since the covering rough set was proposed, there have been two studies directions, approximate operators and multi-scale covering. In [31], Zhang unified the operators of covering rough set in neighborhood systems based on elements, granules and subsystems and then presented the properties. In [22], Yao summarized and generalized the framework of covering approximate operators based on element, granularity and subsystem. Both of the works are basic

but very important. Zhang et al. [29] described the covering rough set in an axiomatic way and gave a more detailed axiom set. Axiomatic characterizations of covering-based approximation operators guarantee the existence of coverings reproducing the operators. In terms of multi-scale covering, Li [7] constructed the scale relation based on covering and extended the multi-scale covering rough set model. Since then, the relation between scales is no longer limited to the usual coarse and fine relationships. The study is also extended to the fuzzy set field. For example, Zhan [25] studied multi-attribute decision making based on the multi-grain (I, T) fuzzy rough set model. Combined with the fuzzy set, theoretical and applied research on rough sets will become more and more extensive.

In 1991, Pawlak began to study the rule extraction in rough set [13]. However, Pawlak's model was only applied to symbolic data. Thus, the extended researches of rule extraction have been made by many scholars. In 1998, Kryszkiewicz proposed an incomplete information system based on tolerance relationship and derived its knowledge [5, 6]. In 2011, Duyong et al. [2] utilized the neighborhood rough set model to extract rule on numerical data. Zhang proposed a rule classification method for symbolic data in the covering [26]. Further, many researchers realized that data structures have complexity and heterogeneity and tried to overcome it [11, 28]. In 2013, Zhang et al. [28] realized multi-confidence rule acquisition in the covering decision system by using the idea of combinatorial optimization. Meng and Shi explored the rule acquisition in incomplete heterogeneous information system in 2020 [11]. But these studies focused more on the attribute reduction rather than on generalization ability of rules. In covering rough set, different description granules of an object may support the same decision in different degrees. In 2017, Wu [18] combined belief function and plausibility function with rough set theory to extract rule in incomplete multi-scale information systems. The work was systematically and completely but didn't show the different effects of the description granules to decision granule. The existing methods do not consider the interactions between different granules, so the extracted rules are poorly adapted. In addition, decision classes in information systems determine the diversity of knowledge, yet little research has been done on this aspect. Motivated by this, we will discuss the new mechanism of rule acquisition based on covering rough sets.

## 3 Preliminaries

Different relations but not limited to equivalence relations yield different granularities. A coverage can be considered as the granularity of a finite and nonempty universe

**Table 1** Outpatient symptom record form (OSRF)

ID	body temperature(°C)	cough	difficult breathing	limbs weakness	Suffering from COVID-19
201	37.0	slight	no	no	no
202	37.4	severe	slight	severe	yes
203	36.8	no	slight	no	no
204	36.8	no	no	general	no
205	36.7	no	no	no	no
206	37.5	slight	general	general	yes
207	37.4	general	no	slight	no
208	37.0	general	general	no	no
209	37.0	slight	general	no	no
210	37.5	severe	severe	severe	yes

of discourse  $U = \{x_1, x_2, \dots, x_n\}$ . In this section, we review some basic conceptual knowledge of covering information system (CIS) and rule extraction. Please see References [14, 16, 22, 32] for more details.

For a decision information system (DIS)  $(U, A, d)$ , define a binary relation  $R_a = \{(x, y) \in U \times U | a(x) = a(y)\}$ , where  $a \in A$  is called a condition attribute,  $a(\cdot)$  represents the attribute value of the object, and the partition of  $U$  consisting of all decision classes is denoted as  $U/d$ . Clearly, we can obtain a covering  $\{(x)_a | x \in U\}$  of  $U$  by  $R_a$ , where  $(x)_a = \{y \in U | (x, y) \in R_a\}$ . This is one but not the only way to generate coverage by a DIS. Thus the CIS is defined as follows.

**Definition 1** A covering information system is an ordered pair  $(U, \mathbb{C})$ , where  $\mathbb{C}$  is a family of coverings of  $U$ . Meanwhile,  $(U, \mathbb{C}, d)$  is known as the covering decision information system, where  $d$  is a decision attribute.

Notice that attribute induces the binary relation on  $U$  to further generate granularity in the covering information system. Generally, complete, incomplete, set-valued or interval-valued information systems can be handled by the covering information system approach. In Example 1, we imitate the application of a covering information system in a medical diagnosis scenario.

**Example 1** In medical diagnosis, giving a diagnosis based on symptoms is making a decision. Table 1 records the symptoms of 10 people who went to the doctor at a hospital and the final diagnosis of COVID-19. Normal body temperature (BT) is less than 37.3°C, otherwise there is a fever. There are four degrees of cough, difficult breathing (DB), and limbs weakness (LW): no, slight, general and severe. In terms of suffering from COVID-19 (SfC), we have two decision classes: no and yes. How do you generate a covering information system based on Table 1?

For the conditional attribute body temperature, group the ID with the same temperature value together, so we get a covering as follows,

$$C_{BT} = \{K_{BT,1}, K_{BT,2}, K_{BT,3}, K_{BT,4}, K_{BT,5}\}.$$

Where  $K_{BT,1} = \{205\}$ ,  $K_{BT,2} = \{203, 204\}$ ,  $K_{BT,3} = \{201, 208, 209\}$ ,  $K_{BT,4} = \{202, 207\}$ , and  $K_{BT,5} = \{206, 210\}$ .

When it comes to the three condition attributes cough, DB and LW, the IDs are clustered together with the same degree of symptoms, respectively. With that there are three coverings here as follows,

$$C_{cough} = \{K_{cough,1}, K_{cough,2}, K_{cough,3}, K_{cough,4}\},$$

$$C_{DB} = \{K_{DB,1}, K_{DB,2}, K_{DB,3}, K_{DB,4}\},$$

$$C_{LW} = \{K_{LW,1}, K_{LW,2}, K_{LW,3}, K_{LW,4}\}.$$

Where  $K_{cough,1} = \{203, 204, 205\}$ ,  $K_{cough,2} = \{201, 206, 209\}$ ,  $K_{cough,3} = \{207, 208\}$ ,  $K_{cough,4} = \{202, 210\}$ ,  $K_{DB,1} = \{201, 204, 205, 207\}$ ,  $K_{DB,2} = \{202, 203\}$ ,  $K_{DB,3} = \{206, 208, 209\}$ ,  $K_{DB,4} = \{210\}$ ,  $K_{LW,1} = \{201, 203, 205, 208, 209\}$ ,  $K_{LW,2} = \{207\}$ ,  $K_{LW,3} = \{204, 206\}$ ,  $K_{LW,4} = \{202, 210\}$ . As a result, we obtain a covering information system  $(U_{OSRF}, \mathbb{C}_{OSRF})$ , where  $U = \{201, 202, \dots, 210\}$ ,  $\mathbb{C}_{OSRF} = \{C_{BT}, C_{cough}, C_{DB}, C_{LW}\}$ .

**Definition 2** A neighborhood operator is a function  $N : U \rightarrow 2^U$ .

Construct a pair of dual element-based lower and upper approximation operators of equivalence class  $D \in U/d$  by means of neighborhood operators as follows,

$$\underline{apr}_N(D) = \{x \in U | N(x) \subseteq D\}, \overline{apr}_N(D) = \{x \in U | N(x) \cap D \neq \emptyset\}.$$

In [22], Yao summarized four common neighborhood operators for a covering  $C$  of  $U$  as follows,

$$\begin{aligned}
N_1(x) &= \bigcap C_x = \bigcap \{K \in C \mid x \in K\}, \\
N_2(x) &= \bigcup C_x = \bigcup \{K \in C \mid x \in K\}, \\
N_3(x) &= \bigcap Md_c\{x\} = \bigcap \{K \in C_x \mid \forall L \in C_x (K \subseteq L \Rightarrow L = K)\}, \\
N_4(x) &= \bigcup md_c\{x\} = \bigcup \{K \in C_x \mid \forall L \in C_x (L \subseteq K \Rightarrow L = K)\}.
\end{aligned}$$

In addition, there are two pairs of dual granule-based lower and upper approximation operators as follows,

$$\begin{aligned}
apr'_C(D) &= \bigcup \{K \in C \mid K \subseteq D\}, \\
\overline{apr}'_C(D) &= U - apr'_C(U - D), \\
apr''_C(D) &= U - \overline{apr}''_C(U - D), \\
\overline{apr}''_C(D) &= \bigcup \{K \in C \mid K \cap D \neq \emptyset\}.
\end{aligned}$$

Rule acquisition is to extract judgment criteria with logical connection from empirical data. The symbols  $\wedge$ ,  $\vee$  and  $\rightarrow$  denote the conjunction, disjunction, and implication, respectively. Any attribute value pair  $(a, v)$  is called atom, where  $a$  is the attribute,  $v \in V_a$ . The logical connection between atoms is called a description. If  $t$  is a description, the attribute appearing in  $t$  is denoted as

$$A(t) = \{a \mid (a, v) \in t\}.$$

The object set with description is called the support set represented by

$$||t|| = \{x \in U \mid a(x) = v, \forall (a, v) \in t\}.$$

The set consisting of all the condition descriptions of the information system is denoted as  $T$ .

**Definition 3** Assume that  $t$  is a description consisting of condition attributes and their attribute values, and  $s = (d, v)$  is a decision description,  $v \in V_d$ . Then  $t \rightarrow s$  is called a decision rule. Sometimes we also use  $r$  to denote a decision rule.

The certain factor of the decision rule is

$$Cer(t \rightarrow s) = \frac{||t \cap s||}{||t||}.$$

The decision rule with  $Cer(t \rightarrow s) = 1$  is called certain rule, and the decision rule with  $0 < Cer(t \rightarrow s) < 1$  is called the possible rule. The validity of the rule can be judged by the certain factor, see Example 2.

**Example 2** According to Example 1, three rules are obtained as follows,

$$t_1 \rightarrow s_1 : (BT, 37.0) \wedge (\text{cough}, \text{slight}) \wedge (DB, \text{no}) \wedge (LW, \text{no}) \rightarrow (SfC, \text{no}),$$

$$\begin{aligned}
t_2 \rightarrow s_1 &: (BT, 37.5) \wedge (\text{cough}, \text{severe}) \wedge (DB, \text{severe}) \wedge (LW, \text{severe}) \rightarrow (SfC, \text{no}), \\
t_3 \rightarrow s_2 &: (BT, 37.5) \wedge (\text{cough}, \text{severe}) \wedge (DB, \text{severe}) \wedge (LW, \text{severe}) \rightarrow (SfC, \text{yes}).
\end{aligned}$$

Calculate their certain factors as follows,

$$\begin{aligned}
Cer(t_1 \rightarrow s_1) &= \frac{|\{201\} \cap \{201, 203, 204, 205, 207, 208, 209\}|}{|\{201\}|} = 1, \\
Cer(t_2 \rightarrow s_1) &= \frac{|\{210\} \cap \{201, 203, 204, 205, 207, 208, 209\}|}{|\{210\}|} = 0, \\
Cer(t_3 \rightarrow s_2) &= \frac{|\{210\} \cap \{202, 206, 210\}|}{|\{210\}|} = 1.
\end{aligned}$$

Since  $Cer(t_2 \rightarrow s_1) = 0$ ,  $t_2 \rightarrow s_1$  is an invalid rule. On the contrary, the other two are valid rules. In addition, we obtain the lower and upper approximations of the two decision classes  $||(\text{SfC}, \text{no})||$  and  $||(\text{SfC}, \text{yes})||$  as follows,

$$\begin{aligned}
apr_{N_1}(|(\text{SfC}, \text{no})|) &= \overline{apr}_{N_1}(|(\text{SfC}, \text{no})|) \\
&= \{201, 203, 204, 205, 207, 208, 209\}, \\
apr_{N_1}(|(\text{SfC}, \text{yes})|) &= \overline{apr}_{N_1}(|(\text{SfC}, \text{yes})|) = \{202, 206, 210\}.
\end{aligned}$$

Furthermore, it is easy to conclude that the lower and upper approximations of equivalence classes generate the certain rules and possible rules respectively.

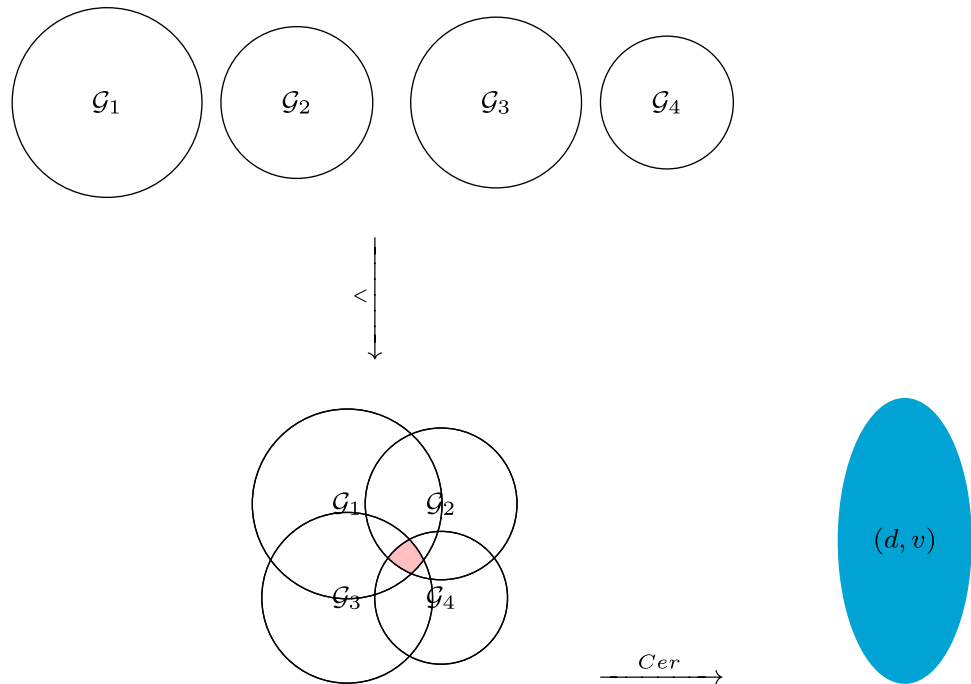
## 4 The mechanism of rule extraction

The classical rough set method theory presents that a conjunctions operation is used to combine different granules to construct the antecedents of rules. Fig. 1 shows the action mechanism of the classical rule extraction method on different granules. In Fig. 1,  $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$  and  $\mathcal{G}_4$  respectively represent granules from different attributes, then the intersection is obtained through intersection operation  $\wedge$ , and finally the corresponding rules are obtained by using the certain factor  $Cer$ . In this section, a new mechanism of rule acquisition in covering information system is first proposed.

### 4.1 Granule associativity rule in covering information system

In a granularity space, granules are the basic unit of computation. People can detect the validity of rules by the interaction between granules. The definition of the rule validity function is presented as follows.

**Definition 4** Let  $(U, \mathbb{C}, d)$  be a granularity decision space, where  $\mathbb{C}$  is a family of condition granularity, and  $d$  is a decision granularity,  $|U| = n, |C| = m$ . Consider a function

**Fig. 1** Rule extraction based on different granules

$V : \prod_{j=1}^m C_j \times d \rightarrow [0, +\infty)$  which we call a rule validity function on  $(U, \mathbb{C}, d)$  if it satisfies the condition that for any  $D'$  and  $D''$ ,  $V(K_1, \dots, K_m, D') \leq V(K_1, \dots, K_m, D'')$ , where  $K_j \in C_j$ ,  $D', D'' \in d$ ,  $D' \subseteq D''$ ,  $j = 1, \dots, m$ . Sometimes denote  $V$  as  $V(t \rightarrow s)$ , where  $||t|| = K_1 \cap \dots \cap K_m$ , and  $||s|| \in d$ .

A larger value for the validity of a rule indicates that the rule is more trustworthy. It is obvious that the certain factor  $Cer$  is a rule validity function. We can conclude from Fig. 1 that in the classical rule extraction method, all the condition granules involved in the rule have the same effect on the decision granule. However, in reality different condition granules act differently on the decision granule which causes discrepancies with reality in the classical approach. As a result we employ granule action parameters to address this problem and then propose granule associativity rule, as follows.

**Definition 5** Given a rule validity function  $V$ , suppose  $\eta_j \geq 0$ , where  $\eta_j$  is the action parameter of  $K_j$  on decision granule  $D$ ,  $j = 1, \dots, m$ . If  $t_j$  is a description of  $K_j$ , then  $t_1 \wedge \dots \wedge t_m \rightarrow s$  is said to be a granule associativity rule with the validity  $V(\eta_1 K_1, \dots, \eta_m K_m, D)$ , where  $s$  is a description of  $D$ .

When  $V = Cer$ , assuming

$$V(\eta_1 K_1, \dots, \eta_m K_m, D) = Cer(t_1 \wedge \dots \wedge t_m \rightarrow s) \cdot \left( \sum_{j=1}^m \eta_j \right) / m, \quad (1)$$

we can use Fig. 2 to illustrate the granule associativity rule extraction process. The color-filled blocks in Fig. 2 indicate that they need to be taken into account when we do rule extraction.

Comparing Fig. 2 with Fig. 1, it can be concluded that the rule extraction method proposed in this paper comprehensively considers the effect of condition granules on decision granules to extract granule associativity rules by (1). In addition, an interesting and significant theorem is obtained as follows.

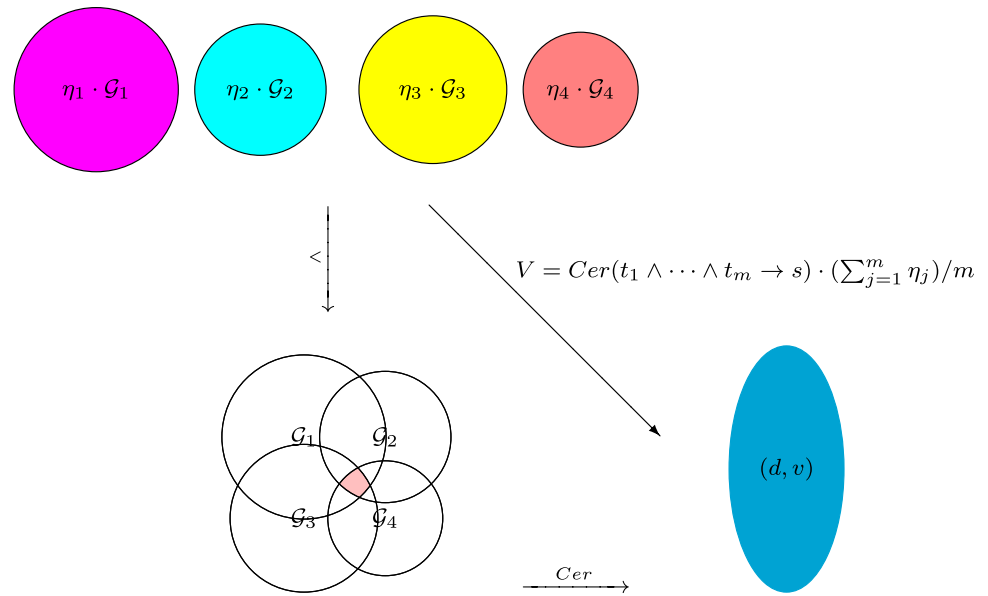
**Theorem 1** When  $\eta_1 = \dots = \eta_m = 1$ , the granule associativity rule is the same as the ordinary rule.

**Proof** It is obvious according to Definition 5.

Table 2, similar to Table 1, records the symptoms of 10 individuals and the final diagnosis of COVID-19. The following example shows the use of granule associativity rule in reality.

**Example 3** According to Table 2, these condition granules can be obtained as follows,

$$\begin{aligned} ||(BT, 36.8)|| &= \{301, 302, 303\} & , \\ ||(BT, 37.3)|| &= \{305, 306, 308, 309\}, \\ ||(BT, 37.5)|| &= \{304, 307, 310\}, \\ ||(cough, no)|| &= \{301, 302, 303\} & , \\ ||(cough, slight)|| &= \{305, 306\}, \\ ||(cough, general)|| &= \{308, 309\} & , \\ ||(cough, severe)|| &= \{304, 307, 310\}, \end{aligned}$$

**Fig. 2** Granule associativity rule based on different granules**Table 2** Outpatient symptom record form

ID	body temperature(°C)	cough	difficult breathing	limbs weakness	Suffering from COVID-19
301	36.8	no	no	no	no
302	36.8	no	no	no	no
303	36.8	no	general	severe	yes
304	37.5	severe	severe	severe	yes
305	37.3	slight	general	severe	yes
306	37.3	slight	general	severe	no
307	37.5	severe	severe	severe	yes
308	37.3	general	general	general	yes
309	37.3	general	general	general	no
310	37.5	severe	severe	severe	no

$|(DB, no)| = \{301, 302\}$ ,  
 $|(DB, general)| = \{303, 305, 306, 308, 309\}$ ,  
 $|(DB, severe)| = \{304, 307, 310\}$ ,  
 $|(LW, no)| = \{301, 302\}$ ,  $|(LW, general)| = \{308, 309\}$ ,  
 $|(LW, severe)| = \{303, 304, 305, 306, 307, 310\}$ .

$|(SfC, no)| = \{301, 302, 306, 309, 310\}$  and  
 $|(SfC, yes)| = \{303, 304, 305, 307, 308\}$  are decision granules.

It is assumed that the novel coronavirus undergoes three periods of outbreak, stabilization and mutation, respectively. The action parameters of condition granules  $(BT, \cdot)$ ,  $(cough, \cdot)$ ,  $(DB, \cdot)$  and  $(LW, \cdot)$  are denoted by  $\eta_1$ ,  $\eta_2$ ,  $\eta_3$  and  $\eta_4$  respectively. During outbreak period (O-P),  $(\eta_1, \eta_2, \eta_3, \eta_4) = (1.5, 1.2, 1.4, 1.1)$ ; during stabilization period (S-P),  $(\eta_1, \eta_2, \eta_3, \eta_4) = (1, 1, 1, 1)$ ; during mutation

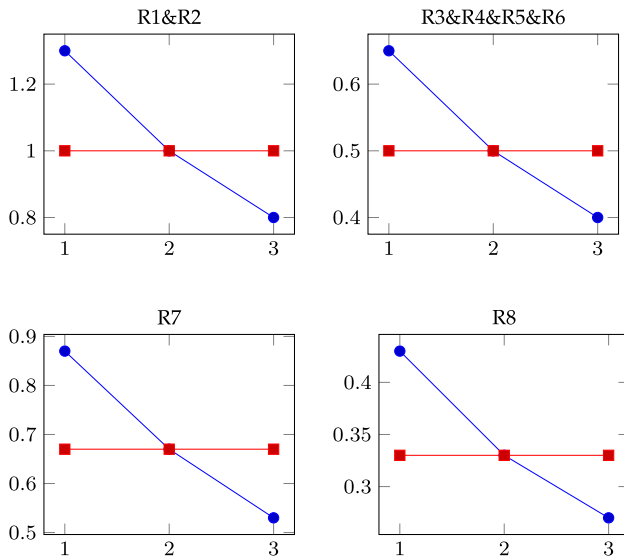
period (M-P),  $(\eta_1, \eta_2, \eta_3, \eta_4) = (0.8, 0.7, 0.9, 0.8)$ . At the same time,  $V = Cer(t_1 \wedge \dots \wedge t_m \rightarrow s) \cdot (\sum_{j=1}^m \eta_j) / m$ . Based on Definitions 4 and 5, the granule associativity rules are extracted as follows, as shown in Table 3.

The columns 3-6 of Table 3, respectively, represent the certain factor, the validity during the outbreak, stabilization and mutation periods. From Table 3, it can be seen that the rule validity is generally minimum during the period of mutation. In addition, the rule validity of R1 and R2 are equal, and the rule validity of R3, R4, R5 and R6 are equal. Therefore, according to the data in Table 3, the variation of the rule validity in different periods can be obtained, as shown in Fig. 3. In Fig. 3, the red line represents the variation of the rule validity using the traditional method for three different periods, while the blue line shows the application of the method proposed in this paper. The abscissa values 1, 2 and 3 indicate the outbreak, stabilization and mutation



**Table 3** The granule associativity rules and validity

ID	Rules	<i>Cer</i>	O-P	S-P	M-P
R1	$(BT,36.8) \wedge (cough,no) \wedge (DB,no) \wedge (LW,no) \rightarrow (SfC,no)$	1	1.3	1	0.8
R2	$(BT,36.8) \wedge (cough,no) \wedge (DB,general) \wedge (LW,severe) \rightarrow (SfC,yes)$	1	1.3	1	0.8
R3	$(BT,37.3) \wedge (cough,slight) \wedge (DB,general) \wedge (LW,severe) \rightarrow (SfC,yes)$	0.5	0.65	0.5	0.4
R4	$(BT,37.3) \wedge (cough,slight) \wedge (DB,general) \wedge (LW,severe) \rightarrow (SfC,no)$	0.5	0.65	0.5	0.4
R5	$(BT,37.3) \wedge (cough,general) \wedge (DB,general) \wedge (LW,general) \rightarrow (SfC,yes)$	0.5	0.65	0.5	0.4
R6	$(BT,37.3) \wedge (cough,general) \wedge (DB,general) \wedge (LW,general) \rightarrow (SfC,no)$	0.5	0.65	0.5	0.4
R7	$(BT,37.5) \wedge (cough,severe) \wedge (DB,severe) \wedge (LW,severe) \rightarrow (SfC,yes)$	0.67	0.87	0.67	0.53
R8	$(BT,37.5) \wedge (cough,severe) \wedge (DB,severe) \wedge (LW,severe) \rightarrow (SfC,no)$	0.33	0.43	0.33	0.27

**Fig. 3** The variation of rule validity during different periods

periods, respectively. Through the analysis of Fig. 3, the conclusions can be drawn as follows.

- The rule validity decreases as the pneumonia virus transitions from outbreak to mutation. Therefore, during the mutation period, we need to find new symptoms to improve rules.
- Traditional method cannot reflect the variation of rule validity in different periods, but the method proposed in this paper can.

## 4.2 Rule-based data potential

In a covering information system, the ultimate diversity of rules lies in the decision classes. The identification of decision class, which is one of the most important applications of information system, depends on the object in  $U$ . Inspired by the rules, the rule-based classification of object is carried out.

**Definition 6** Given a rule validity function  $V$ , let  $s$  be a decision description,  $\alpha$  and  $\beta$  are two non-negative parameters,  $\alpha \geq \beta$ . For  $x \in U$ ,  $T_x = \{t \in T | x \in ||t||\}$ ,

1. if for any  $t \in T_x$ ,  $V(t \rightarrow s) \geq \alpha$ , then write  $x \Rightarrow_{\geq \alpha} s$ ;
2. if there exists  $t_1, t_2 \in T_x$ ,  $V(t_1 \rightarrow s) < \alpha$ , and  $V(t_2 \rightarrow s) > \beta$ , then write  $x \Rightarrow_{(\beta, \alpha)} s$ ;
3. if for any  $t \in T_x$ ,  $V(t \rightarrow s) \leq \beta$ , then write  $x \Rightarrow_{\leq \beta} s$ .

According to Definition 6, when  $x \Rightarrow_{\geq \alpha} s$ ,  $x$  strongly supports decision class  $||s||$ . When  $x \Rightarrow_{(\beta, \alpha)} s$ ,  $x$  generally supports decision class  $||s||$ . However, when  $x \Rightarrow_{\leq \beta} s$ , the decision class  $||s||$  is barely supported by  $x$ . The potential of an information system for a decision class is determined by the object. To measure the potential, the sets  $\{x \in U | x \Rightarrow_{\geq \alpha} s\}$ ,  $\{x \in U | x \Rightarrow_{(\beta, \alpha)} s\}$  and  $\{x \in U | x \Rightarrow_{\leq \beta} s\}$  are denoted as  $\Omega_1^s$ ,  $\Omega_2^s$  and  $\Omega_3^s$ , respectively. Meanwhile,  $\theta_1 = |\Omega_1^s|$ ,  $\theta_2 = |\Omega_2^s|$ , and  $\theta_3 = |\Omega_3^s|$ . Obviously  $\Omega_1^s$ ,  $\Omega_2^s$  and  $\Omega_3^s$  constitute a partition of  $U$ .

**Definition 7** Assume that the variable  $E \geq 0$ . If  $E$  is respectively positive correlation with  $\theta_1$  and  $\theta_2$ , and negative correlation with  $\theta_3$ , then  $E$  is said to be rule-based data potential for  $s$ .

In accordance with Definition 7, we attempt to construct the rule-based data potential calculation formula as follows,

$$E = \frac{\exp(\theta_1) + \theta_2 - 1}{\theta_3 + 1} \quad (2)$$

Study of Equation (2) reveals that when  $\theta_1$  and  $\theta_2$  are increased by the same amount separately, the increase brought by  $\theta_1$  of  $E$  is greater than the increase brought by  $\theta_2$  of  $E$ . On the other hand when  $\theta_1 = \theta_2 = 0$ ,  $E = 0$ . The denominator  $\theta_3 + 1$  is not going to be equal to 0.

In the case of  $\eta_1 = \dots = \eta_m = 1$  and  $\alpha = 1$ ,  $\beta = 0$ , the rule extraction can be transformed into classical rule extraction by Equation (1). An interesting theorem is presented as follows.

**Table 4** The performance and salary of the employees

ID	CW	NoA	The level of salary
1001	100	4	I
1002	400	2	II
1003	600	1	III
1004	200	3	I
1005	200	3	II
1006	100	4	I
1007	100	3	I
1008	300	2	II
1009	400	2	II
1010	500	1	III

**Theorem 2** Based on Equation (1), if  $\eta_1 = \dots = \eta_m = 1$  and  $\alpha = 1, \beta = 0$ , then hold that for a covering decision information system  $(U, \mathbb{C}, d)$ ,

$$\begin{aligned} \Omega_2^s &= \{x \in U \mid \exists t_1, t_2 \in T_x, \|t_1\| \cap \|s\| \neq \emptyset / \\ &= \|t_1\| \cap \|t_2\| \cap \|s\| \neq \emptyset \mid \forall t \in T_x, \|t\| \subseteq \|s\|\} \\ 1. &= \{x \in U \mid (\forall C \in \mathbb{C}, \|C\| \cap \|s\| \neq \emptyset) \Rightarrow \|x\| \subseteq \|s\|\} \\ 2. &\Rightarrow (K \cap D \neq \emptyset \wedge \|x\| \cap \|s\| \neq \emptyset) \Rightarrow \|x\| \subseteq \|s\|, \\ 3. &= \{x \in U \mid (\forall C \in \mathbb{C}) \wedge (\forall K \in \mathbb{K}) \Rightarrow K \cap D = \emptyset\}, \\ 4. &\Omega_1^s \subseteq \overline{\text{apr}}_{N_1}(\|s\|) \subseteq \overline{\text{apr}}_{N_1}(\|s\|) \subseteq \Omega_1^s \cup \Omega_2^s, \end{aligned}$$

where  $D = \|s\|$ .

**Proof** According to Equation (1), since  $V = \text{Cer}$ , we have  $V(t \rightarrow s) = \frac{\|t\| \cap \|s\|}{\|t\|}$ . Further, we can get  $V(t \rightarrow s) = 1 \Leftrightarrow \|t\| \subseteq \|s\|$  and  $V(t \rightarrow s) = 0 \Leftrightarrow \|t\| \cap \|s\| = \emptyset$ . In addition, a one-to-one mapping can be established between  $T$  and  $\mathbb{C}$ . Hence, based on Definition 6, Theorem 2 is straightforward.

The traditional method of rule extraction lacks the evaluation of the information regarding decision class, which leads to the inability to pick a suitable information system. Theorem 2 shows that the rule-based data potential can be considered in classical rule extraction. Item (4) of Theorem 2 states that based on  $\Omega_1^s$  and  $\Omega_2^s$ , the approximation accuracy of the equivalence class is not reduced. A greater data potential refers to the more information contained in the information system. As a result, after finding the specific rule validity and rule-based data potential,  $E$  can be employed as the measurement criterion for choosing an information system. In other words, for a decision description  $(d, v)$ , an information system with a greater  $E$  should be selected for consideration. On the other hand, based on rule-based data potential, a comprehensive comparison of all decision descriptions can be drawn. An example of the application of rule-based data potential in business management is given as follows.

**Example 4** Wages are the key to good management. There are two groups of employees in a certain department of a company. The two groups of employees are numbered 1001 to 1010, and 2001 to 2010, respectively. Tables 4 and 5 collect the performance and salary of the employees respectively. There are three levels of salary, I:1000~1100, II:1100~1500 and III:1500~2000, unit: US dollar. Two factors that affect salary are the completed workload (CW) and the number of absences (NoA). The company needs to choose either Tables 4 or 5 as the reference for the coming salary level.

From Table 4, we derive that as follows,

$$\begin{aligned} C'_{CW,1} &= C'_{CW,6} = C'_{CW,7} = \{\{1001, 1006, 1007\}\}, \\ C'_{CW,2} &= C'_{CW,9} = \{\{1002, 1009\}\}, \\ C'_{CW,3} &= \{\{1003\}\}, \\ C'_{CW,4} &= C'_{CW,5} = \{\{1004, 1005\}\}, \\ C'_{CW,8} &= \{\{1008\}\}, \\ C'_{CW,10} &= \{\{1010\}\}, \\ C'_{NoA,1} &= C'_{NoA,6} = \{\{1001, 1006\}\}, \\ C'_{NoA,2} &= C'_{NoA,8} = C'_{NoA,9} = \{\{1002, 1008, 1009\}\}, \\ C'_{NoA,3} &= C'_{NoA,10} = \{\{1003, 1010\}\}, \\ C'_{NoA,4} &= C'_{NoA,5} = C'_{NoA,7} = \{\{1004, 1005, 1007\}\}, \\ D'_I &= \{1001, 1004, 1006, 1007\}, \\ D'_{II} &= \{1002, 1005, 1008, 1009\}, \\ D'_{III} &= \{1003, 1010\}. \end{aligned}$$

From Table 5, we derive that as follows,

$$\begin{aligned} C''_{CW,1} &= \{\{2001\}\}, \\ C''_{CW,2} &= C''_{CW,8} = \{\{2002, 2008\}\}, \\ C''_{CW,3} &= C''_{CW,6} = C''_{CW,7} = \{\{2003, 2006, 2007\}\}, \\ C''_{CW,4} &= \{\{2004\}\}, \\ C''_{CW,5} &= C''_{CW,9} = \{\{2005, 2009\}\}, \\ C''_{CW,10} &= \{\{2010\}\}, \\ C''_{NoA,1} &= C''_{NoA,10} = \{\{2001, 2010\}\}, \\ C''_{NoA,2} &= C''_{NoA,8} = \{\{2002, 2008\}\}, \\ C''_{NoA,3} &= C''_{NoA,4} = C''_{NoA,6} = C''_{NoA,7} = \{\{2003, 2004, 2006, 2007\}\}, \end{aligned}$$

**Table 5** The performance and salary of the employees

ID	CW	NoA	The level of salary
2001	500	1	III
2002	100	4	I
2003	200	3	I
2004	300	3	II
2005	400	2	II
2006	200	3	I
2007	200	3	II
2008	100	4	I
2009	400	2	II
2010	600	1	III



$$\begin{aligned}C''_{NoA,5} &= C''_{NoA,9} = \{\{2005, 2009\}\}, \\D''_I &= \{2002, 2003, 2006, 2008\}, \\D''_{II} &= \{2004, 2005, 2007, 2009\}, \\D''_{III} &= \{2001, 2010\}.\end{aligned}$$

According to Theorem 2, the results are displayed as follows by calculation.

For Table 4,

$$\begin{aligned}\Omega^I_1 &= \{1001, 1006\}, \Omega^I_2 = \{1004, 1005, 1007\}, \\ \Omega^I_3 &= \{1002, 1003, 1008, 1009, 1010\}, \\ \Omega^{II}_1 &= \{1002, 1008, 1009\}, \Omega^{II}_2 = \{1004, 1005, 1007\}, \\ \Omega^{II}_3 &= \{1001, 1003, 1006, 1010\}, \\ \Omega^{III}_1 &= \{1003, 1010\}, \Omega^{III}_2 = \emptyset, \\ \Omega^{III}_3 &= \{1001, 1002, 1004, 1005, 1006, 1007, 1008, 1009\}.\end{aligned}$$

For Table 5,

$$\begin{aligned}\Omega^I_1 &= \{2002, 2008\}, \Omega^I_2 = \{2003, 2004, 2006, 2007\}, \\ \Omega^I_3 &= \{2001, 2005, 2009, 2010\}, \\ \Omega^{II}_1 &= \{2005, 2009\}, \Omega^{II}_2 = \{2003, 2004, 2006, 2007\}, \\ \Omega^{II}_3 &= \{2001, 2002, 2008, 2010\}, \\ \Omega^{III}_1 &= \{2001, 2010\}, \Omega^{III}_2 = \emptyset, \\ \Omega^{III}_3 &= \{2002, 2003, 2004, 2005, 2006, 2007, 2008, 1009\}.\end{aligned}$$

Based on Equation (2), we can obtain that

$$\begin{aligned}E_{I,Table4} &= \frac{exp(2)+2}{5}, E_{II,Table4} = \frac{exp(3)+2}{5}, E_{III,Table4} = \frac{exp(2)-1}{9}, \\ E_{I,Table5} &= \frac{exp(2)+3}{5}, E_{II,Table5} = \frac{exp(2)+3}{5}, E_{III,Table5} = \frac{exp(2)-1}{9}.\end{aligned}$$

Since  $E_{I,Table5} > E_{I,Table4}$ , choose Table 5 as the reference for the salary level I; since  $E_{II,Table4} > E_{II,Table5}$ , choose Table 4 as the reference for the salary level II; since  $E_{III,Table4} = E_{III,Table5}$ , choose Tables 4 or 5 as the reference for the salary level III.

## 5 Scale selection based on rule extraction

Scale selection can effectively improve the generalization ability of the covering information system. However, the different scales inherently cause a shift in the rules. In this section, scale selection based on rule extraction is explored on multi-scale covering information system.

**Definition 8** Let  $C'$  and  $C''$  be two coverings of  $U$ , respectively. We say that there is a scale relationship between  $C'$  and  $C''$ , denoted as  $C' \triangleright C''$ , if and only if there exists a function  $f : C' \rightarrow 2^{C''}$  satisfying the following conditions:

1.  $\forall K'' \in C'', \exists K \in 2^{C'}, \text{ s.t. } K'' = \bigcup K, \forall K \in \mathbb{K}, K'' \in f(K),$
2.  $\forall K_1 \in C', K_2 \in f(K_1), \text{ s.t. } K_1 \subseteq K_2,$
3.  $\bigcup_{K \in C'} f(K) = C''.$

A covering information system  $(U, \mathbb{C}_1 \cup \dots \cup \mathbb{C}_m)$  is said to be a multi-scale covering information system if for any  $C', C'' \in \mathbb{C}_j, C' \triangleright C''$ , where  $j = 1, \dots, m$ ,  $m$  is a positive integer, and  $\mathbb{C}_j$  is a family of coverings. At the same time,  $(U, \mathbb{C}_1 \cup \dots \cup \mathbb{C}_m, d)$  is a multi-scale covering decision information system, where  $d$  is a partition of  $U$  composed of equivalence classes. There are  $I_1$  scales in  $\mathbb{C}_1, I_2$  scales in  $\mathbb{C}_2, \dots, I_m$  scales in  $\mathbb{C}_m$ . The set of scale combinations is denoted as  $\mathbb{S}$ , i.e.  $\mathbb{S} = \{C_1^{j_1}, \dots, C_m^{j_m}\}$ , where  $1 \leq j_1 \leq I_1, \dots, 1 \leq j_m \leq I_m$ .  $C_1^{j_1}$  is induced by the  $j_1$ th scale of  $a_1, C_m^{j_m}$  is induced by the  $j_m$ th scale of  $a_m$  and so on. Moreover,  $S_1 = \{C_1^1, \dots, C_m^1\}$  represents the first scale combination. The relation between the different scales of the covering family induced by the  $a_j$  is  $C_j^1 \triangleright C_j^j, 1 \leq j = 1, \dots, m$ . Define a relation ' $\leq$ ', for any  $S', S'' \in \mathbb{S}, S' \leq S''$  if and only if for any  $C_j' \in S', C_j'' \in S'', C_j' \triangleright C_j''$ . Assume that if  $S' \leq S''$  and  $S'' \leq S'$ , then  $S' = S''$ . Thus,  $(\mathbb{S}, \leq)$  is a partial order set.

**Definition 9** In the partial order set  $(\mathbb{S}, \leq)$ ,  $S^*$  is known as optimal scale combination based on rule extraction meeting the requirements as follows,

1.  $\forall s \in U/d, S \in \mathbb{S}, \text{ s.t. } E_s^* \geq E_s,$
2.  $\forall S \in \mathbb{S}, S^* \leq S, \exists s_0 \in U/d, \text{ s.t. } E_{s_0} < E_{s_0}^*,$

where  $E_s^*$  and  $E_{s_0}^*$  are the rule-based data potentials of  $S^*$  for  $s$  and  $s_0$  respectively, and  $E_s, E_{s_0}$  the rule-based data potentials of  $S$  for  $s, s_0$ .

Sometimes, there is no optimal scale combination based on rule extraction in a multi-scale covering decision information system. However, an interesting theorem is presented as follows.

**Theorem 3** In the partial order set  $(\mathbb{S}, \leq)$ , for  $s \in U/d$ , there exists  $S^{*,s} \in \mathbb{S}$  such that

1.  $\forall S \in \mathbb{S}, E_s^* \geq E_s,$
2.  $\forall S \in \mathbb{S}, S^* \leq S \Rightarrow E_s^* > E_s,$

where  $E_s^*$  and  $E_s$  are the rule-based data potentials of  $S^{*,s}$  and  $S$  for  $s$  respectively.

**Proof** Due to the finite set  $\mathbb{S}$ , it is straightforward.

Theorem 3 states that for a decision description in a multi-scale covering decision information system, there is an optimal scale combination that maximizes the data potential

of the description. In the previous method of scale combination selection, the scale combination with greater generalization ability is selected by keeping the lower approximation of the equivalence class of  $S_1$ . What is amazing is the introduction of the following theorem.

**Theorem 4** *Given a first scale combination  $S_1 = \{C_1^1, \dots, C_m^1\}$ , a decision description  $s$ ,  $\underline{apr}_{N_1}^{S_1}(\|s\|)$  and  $\Omega_1^s, \Omega_2^s, \Omega_3^s$ . For  $x \in (\Omega_1^s \cup \Omega_2^s) - \underline{apr}_{N_1}^{S_1}(\|s\|)$ ,  $y \in \Omega_3^s$ ,  $K_x \in C_{j,x}^1$ ,  $K_y \in C_{j,y}^1$ , let  $K_{xy} = K_x \cup K_y$ ,  $C_j = (C_j^1 - \{K_x, K_y\}) \cup \{K_{xy}\}$ ,  $S = \{C_1^1, \dots, C_{j-1}^1, C_j, C_{j+1}^1, \dots, C_m^1\}$ , then it is concluded that*

1.  $S_1 \triangleright S$ ,
2.  $E_s \geq E_s^1$ ,
3.  $\underline{apr}_{N_1}^S(\|s\|) = \underline{apr}_{N_1}^{S_1}(\|s\|)$ ,

where  $j \in \{1, \dots, m\}$ ,  $\underline{apr}_{N_1}^S(\|s\|)$  and  $\underline{apr}_{N_1}^{S_1}(\|s\|)$  are the lower approximations of  $\|s\|$  in  $S$  and  $S_1$ , respectively.

**Proof** According to Theorem 2 and Definition 8, Theorem 4 is straightforward.

Theorem 4 delivers a way to construct optimal scale combination based on rule extraction with keeping the lower approximation. Meanwhile, Theorem 4 theoretically proves that there exists a scale combination such that the

**Table 6** A description of data sets

Data sets <sup>1</sup>	Instances	Condition attributes	Classes
Iris Plants	150	4	3
Hayes-Roth	132	4	3
Somerville Happiness Survey	143	6	2

<sup>1</sup> The data sets do not contain default values

generalization ability and data potential of the model are optimal and the lower approximation is maintained.

## 6 Numerical experiments

In this section, firstly, Algorithm 1 is designed to calculate the data potential. Then experiments are carried out to illustrate the effectiveness of the algorithm and the comparison of data potentials at different scales.

### 6.1 Algorithm

Algorithm 1 is designed to calculate the rule-based data potential for an equivalence class  $D$  of an information

**Table 7** Attribute information in the data sets

Attribute	Value range
sepal length <sup>1</sup>	[4.3, 7.9]
sepal width <sup>1</sup>	[2.0, 4.4]
petal length <sup>1</sup>	[1.0, 6.9]
petal width <sup>1</sup>	[0.1, 2.5]
class <sup>1</sup>	{Iris Setosa, Iris Versicolour, Iris Virginica}({1, 2, 3})
hobby <sup>2</sup>	{1, 2, 3}
age <sup>2</sup>	{1, 2, 3, 4}
educational level <sup>2</sup>	{1, 2, 3, 4}
marital status <sup>2</sup>	{1, 2, 3, 4}
class <sup>2</sup>	{1, 2, 3}
the availability of information about the city services <sup>3</sup>	{1, 2, 3, 4, 5}
the cost of housing <sup>3</sup>	{1, 2, 3, 4, 5}
the overall quality of public schools <sup>3</sup>	{1, 2, 3, 4, 5}
your trust in the local police <sup>3</sup>	{1, 2, 3, 4, 5}
the maintenance of streets and sidewalks <sup>3</sup>	{1, 2, 3, 4, 5}
the availability of social community events <sup>3</sup>	{1, 2, 3, 4, 5}
decision attribute <sup>3</sup>	{happy, unhappy}({1, 2})

<sup>1</sup> From Iris Plants

<sup>2</sup> From Hayes-Roth

<sup>3</sup> From Somerville Happiness Survey

system. In Algorithm 1, if  $p = \prod_{i=1}^m |C_i|$ , then  $x \in \Omega_1$ ; if  $q = \prod_{i=1}^m |C_i|$ , then  $x \in \Omega_3$ . Lastly,  $\Omega_2 = U - \Omega_1 - \Omega_3$ . Let  $h = \max\{|C_1|, \dots, |C_m|\}$ . Lines 4-8 possess the time

complexity  $O(nm)$ . The time complexity of Lines 9-31 is  $O(nmh)$ . Therefore, we can conclude that the time complexity of Algorithm 1 is  $O(nm(h+1))$ .

---

**Algorithm 1:** Calculation of rule-based data potential.

---

**Input:** a covering decision information system  $S = (U, \mathbb{C}, d)$  and the family of coverings  $\mathbb{C} = \{C_1, C_2, \dots, C_m\}$ , an equivalence class  $D$ , a rule validity function  $V$ , two non-negative real numbers  $\alpha$  and  $\beta$ .

**Output:** Rule-based data potential  $E$ .

```

1   $n = |U|, m = |\mathbb{C}|$ ; //  $n$  and  $m$ ,  $w$  are the number of elements in  $U$  and  $\mathbb{C}$ , respectively.
2   $E = 0$ ; // Initialize Rule-based data potential.
3   $O_1 = 0, O_2 = 0, O_3 = 0$ ; // Initialize  $|\Omega_1|, |\Omega_2|$  and  $|\Omega_3|$ .
4  for  $x$  in  $U$  do
5      for  $C \in \mathbb{C}$  do
6          calculate  $C_x$ ; // Calculate neighborhood systems.
7      end
8  end
9  for  $x \in U$  do
10      $p = 0, q = 0$ ; //  $p$  and  $q$  represent the number of granules, respectively.
11     for  $C' \in \mathbb{C}$  do
12         for  $K \in C_x$  do
13             if  $V(K, D) \geq \alpha$  then
14                  $p = p + 1$ ;
15             else
16                 if  $V(K, D) \leq \beta$  then
17                      $q = q + 1$ ;
18                 end
19             end
20         end
21     end
22     if  $p = \prod_{i=1}^m |C_i|$  then
23          $O_1 = O_1 + 1$ ;
24     else
25         if  $q = \prod_{i=1}^m |C_i|$  then
26              $O_3 = O_3 + 1$ ;
27         else
28              $O_2 = O_2 + 1$ ;
29         end
30     end
31 end
32  $E = \frac{\exp(O_1) + O_2 - 1}{O_3 + 1}$ ;
33 return  $E$ .
```

---

## 6.2 Experimental analysis

In order to demonstrate the effectiveness of the algorithm and the differences in data potential at different scales, we randomly obtained three data sets from UCI[3] as the original data of the experiment that fit the rule extraction model, i.e. all have only one decision attribute. All of the data sets are numerical data. These three data sets are 'Iris Plants', 'Hayes-Roth' and 'Somerville Happiness Survey'. Where 'Iris Plants' was also employed by references [11] and [26]. Tables 6 and 7 contain the basic and key information for the data sets.

We directly use the raw data as the data combination at the first scale. By dividing the range equally, we can construct different scale combinations. Table 8 is the basic information of the constructed scale combination. After

combining the attribution, Iris Plants has three scales, Hayes-Roth has two scales, so does Somerville Happiness Survey. Let  $V = \text{Cer}$ ,  $\alpha = 1$ , and  $\beta = 0$ . According to Algorithm 1, MATLAB software is used to calculate the data potential. Some results are shown in Tables 9, 10 and 11. Based on these results, the effectiveness of Algorithm 1 is obvious.

According to Tables 9, 10 and 11, we can get the comparison results of the data potential of information system at different scales. As for 'Iris Plants', it is noted from Table 9 that the original information system has the smallest data potential 5.069 for class 1 in the first scale. In Table 9, as the scale changes from the first to the third,  $|\Omega_2|$  of all equivalence classes is getting

**Table 8** Value range of the attributes

Attribute	The second scale combination	The third scale combination
sepal length	{1, 2, 3, 4, 5}	{1, 2, 3}
sepal width	{1, 2, 3, 4, 5}	{1, 2, 3}
petal length	{1, 2, 3, 4, 5}	{1, 2, 3}
petal width	{1, 2, 3, 4, 5}	{1, 2, 3}
hobby	{1, 2, 3}	*
age	{1, 2, 3}	*
educational level	{1, 2, 3}	*
marital status	{1, 2, 3}	*
the availability of information about the city services	{1, 2, 3}	*
the cost of housing	{1, 2, 3}	*
the overall quality of public schools	{1, 2, 3}	*
your trust in the local police	{1, 2, 3}	*
the maintenance of streets and sidewalks	{1, 2, 3}	*
the availability of social community events	{1, 2, 3}	*

\* means null value

**Table 9** The rule-based data potential of 'Iris Plants' under the different scales

Class	The first scale				The second scale				The third scale			
	$ \Omega_1 $	$ \Omega_2 $	$ \Omega_3 $	$E$	$ \Omega_1 $	$ \Omega_2 $	$ \Omega_3 $	$E$	$ \Omega_1 $	$ \Omega_2 $	$ \Omega_3 $	$E$
1	1	125	24	5.069	0	150	0	150	0	150	0	150
2	0	145	5	24.167	0	148	2	49.333	0	150	0	150
3	0	131	19	6.55	0	150	0	150	0	150	0	150

larger, while  $|\Omega_1|$  and  $|\Omega_3|$  are getting smaller, and thus the data potential is getting larger. However, for 'Hayes-Roth' and 'Somerville Happiness Survey', it can be concluded from Tables 10 and 11 that the scale change does not cause a change in the data potential. In fact, for  $|\Omega_3|$  of all the equivalence classes, their values are 0

in Tables 10 and 11, so it is impossible to construct the scale combination with greater data potential through Theorem 4.

## 7 Conclusion

Compared with the traditional rule extraction model, the granule associativity rule proposed in this paper enjoys better adaptability. The introduced action parameters address the validity of the rule. The objects are classified by rules, and then the measure of covering information systems about the decision class are carried out. The rule-based data potential is presented as a criterion for evaluating information system and is well employed in the scale selection for multi-scale covering rough set. Such a mechanism of rule acquisition further develops the rough set theory.

In the future we will devote to the association of action parameters with fuzzy measures and the use of improved rule extraction methods in larger data scenarios. What's more, the fusion of the three-way decision with the rule extraction mechanism arouses our interest. The theory is advised for pattern recognition, medical diagnosis and other fields.

**Table 10** The rule-based data potential of 'Hayes-Roth' under the different scales

Class	The first scale				The second scale			
	$ \Omega_1 $	$ \Omega_2 $	$ \Omega_3 $	$E$	$ \Omega_1 $	$ \Omega_2 $	$ \Omega_3 $	$E$
1	0	132	0	132	0	132	0	132
2	0	132	0	132	0	132	0	132
3	0	132	0	132	0	132	0	132

**Table 11** The rule-based data potential of 'Somerville Happiness Survey' under the different scales

Class	The first scale				The second scale			
	$ \Omega_1 $	$ \Omega_2 $	$ \Omega_3 $	$E$	$ \Omega_1 $	$ \Omega_2 $	$ \Omega_3 $	$E$
1	0	143	0	143	0	143	0	143
2	0	143	0	143	0	143	0	143

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (No:11871259), Natural Science Foundation of Fujian (No:2019J01748) and the Key Program of the Natural Science Foundation of Fujian Province (No: 2020J02043).

## References

- Chen D, Wang C, Hu Q (2007) A new approach to attribute reduction of consistent and inconsistent covering decision systems with covering rough sets. *Information Sciences* 177:3500–3518
- Du Y, Hu Q, Zhu P, Ma P (2011) Rule learning for classification based on neighborhood covering reduction. *Information Sciences* 181:5457–5467
- Dua D, Graff C (2017) UCI machine learning repository. <http://archive.ics.uci.edu/ml>
- Jensen R, Qiang S (2004) Fuzzy-rough attributes reduction with application to web categorization. *Fuzzy Sets and Systems* 141:469–485
- Kryszkiewicz M (1998) Rough set approach to incomplete information systems. *Information Sciences* 112:39–49
- Kryszkiewicz M (1999) Rules in incomplete information systems. *Information Sciences* 113:271–292
- Li W, Li J, Huang J, Dai W, Zhang X (2021) A new rough set model based on multi-scale covering. *International Journal of Machine Learning and Cybernetics* 12:243–256
- Liu C, Miao D (2011) Covering rough set model based on multi-granulations. *Fuzzy Set, Data Mining and Granular Computing* 6743:87–90
- Liu C, Wang M (2011) Covering fuzzy rough set based on multi-granulations. *Proceedings of International Conference on Uncertainty Reasoning and Knowledge Engineering* 2:146–149
- Liu C, Miao D, Qian J (2014) On multi-granulation covering rough sets. *International Journal of Approximate Reasoning* 55:1404–1418
- Meng Z, Shi Z (2020) On rule acquisition methods for data classification in heterogeneous incomplete decision system. *Knowledge-Based Systems* 193:105472
- Pawlak Z (1982) Rough sets. *International Journal of Computer and Information Sciences* 11:341–356
- Pawlak Z (1991) *Rough Set*. Kluwer Academic Publishers, Theoretical Aspects of Reasoning about Data
- Shi Z, Gong Z (2010) The further investigation of covering-based rough sets: uncertainty characterization, similarity measure and generalized models. *Information Sciences* 180:3745–3763
- Wang J, Dai D, Zhou Z (2004) Fuzzy covering generalized rough sets (in chinese). *Fuzzy Sets and Systems* 2:20–22
- Wang L, Yang X, Yang J, Wu C (2012) Relationships among generalized rough sets in six coverings and pure reflexive neighborhood system. *Information Sciences* 201:66–78
- Wu W, Leung Y (2011) Theory and applications of granular labelled partitions in multi-scale decision tables. *Information Sciences* 181:3878–3897
- Wu W, Qian Y, Li T, Gu S (2017) On rule acquisition in incomplete multi-scale decision tables. *Information Sciences* 378:282–302
- Xu J, Sun L, Zhang Q (2003) *Granular computing and theory method of measurement of uncertain information*(in Chinese). The Science Publishing Company, Beijing
- Yao Y (2001) On generalizing rough set theory. *Proceedings of the Ninth International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing* 2639:44–51
- Yao Y (2006) Neighborhood systems and approximation retrieval. *Information Sciences* 176:3431–3452
- Yao Y, Yao B (2012) Covering based rough set approximations. *Information Sciences* 200:91–107
- Zadeh L (1997) Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems* 90:111–127
- Zakowski W (1983) Approximations in the space  $(u, p)$ . *Demonstratio Mathematica* 16:761–769
- Zhan J, Sun B, Alcantud JCR (2019) Covering based multi-granulation  $(i, t)$ -fuzzy rough set models and applications in multi-attribute group decision-making. *Information Sciences* 476:290–318
- Zhang B, Min F, Ciucc D (2015) Representative-based classification through covering-based neighborhood rough sets. *Applied Intelligence* 43:840–854
- Zhang Q, Wang G, Hu J (2003) *Multi-granularity knowledge acquisition and uncertainty measure* (in Chinese). The Science Publishing Company, Beijing
- Zhang X, Mei C, Chen D, Li J, Huang J, Kong J (2013) Multi-confidence rule acquisition oriented attribute reduction of covering decision systems via combinatorial optimization. *Knowledge-Based Systems* 50:187–197
- Zhang Y, Luo M (2011) On minimization of axiom sets characterizing covering-based approximation operators. *Information Sciences* 181:3032–3042
- Zhang Y, Luo M (2013) Relationships between covering-based rough sets and relation-based rough sets. *Information Sciences* 225:55–71
- Zhang Y, Li C, Lin M, Lin Y (2015) Relationships between generalized rough sets based on covering and reflexive neighborhood system. *Information Sciences* 319:56–67
- Zhu P (2011) Covering rough sets based on neighborhoods: an approach without using neighborhoods. *International Journal of Approximate Reasoning* 52:461–472