ATTRIBUTE SIGNIFICANCE, CONSISTENCY MEASURE AND ATTRIBUTE REDUCTION IN FORMAL CONCEPT ANALYSIS

*C. Huang**, *J. Li**, *S.M. Dias*†

Abstract: One focus of data analysis in formal concept analysis is attribute-significance measure, and another is attribute reduction. From the perspective of information granules, we propose information entropy in formal contexts and conditional information entropy in formal decision contexts, and they are further used to measure attribute significance. Moreover, an approach is presented to measure the consistency of a formal decision context in preparation for calculating reducts. Finally, heuristic ideas are integrated with reduction technique to achieve the task of calculating reducts of an inconsistent data set.

Key words: *formal concept analysis, information entropy, attribute significance, consistency*

Received: April 19, 2016
Revised and accepted: October 3, 2016
DOI: 10.14311/NNW.2016.26.035

1. Introduction

Formal concept analysis (FCA) [39] tries to mine knowledge from a formal context (O, A, R). This kind of knowledge is a special structure called concept lattice that is constituted by formal concepts [39]. Up till now, FCA has gained applications which are knowledge discovery [3, 14, 23, 29], information retrieval [5], machine learning [11], cognitive learning [13, 17], software engineering [26, 28], and so on [31, 35, 41, 45].

Granular computing is a good theory to deal with problems by using the idea of granulation [42–44]. Recently, there have been some researches on granular computing approach of FCA. For example, Ma et al. [20] discussed the relationship between Galois connection and granular computing. Qiu et al. [24] established a concept granular computing system so as to contain as many types of concept lattices as possible in the same mathematical model. Wu et al. [40] examined the granular structures of concept lattices and demonstrated their application in attribute reduction. Considering that entropy theory can be used to solve problems

*Chenchen Huang, Jinhai Li -— Corresponding author, Faculty of Science, Kunming University of Science and Technology, Kunming, Yunnan, PR China, E-mail: jhlixjtu@163.com, huangchenchen64@163.com,

†Sérgio M. Dias, Federal Service of Data Processing (SERPRO), Belo Horizonte, Minas Gerais, Brazil, E-mail: sergio.dias@serpro.gov.br

of measuring attribute significance [19,22,36] and constructing concept lattices [12], we propose in this paper the notions of information entropy in formal contexts and conditional information entropy to solve the problem of measuring attribute significance in FCA.

Attribute reduction in FCA is a classical topic and many researchers tried to investigate it (see e.g. [2,7–9,15,18,21,25,27,30,37,38,47]) . For instance, Dias and Vieira [6] concerned how to divide the existing reduction techniques into different classes so as to make progress in this topic. Particularly, Wu et al. [40] proposed a granular computing based technique to avoid the data redundancy, and pointed out that their reduction methods are computationally expensive due to Boolean reasoning. Furthermore, they claimed that heuristic ideas are required to be integrated with reduction technique to enhance the efficiency of computing granular reducts, and it is still necessary to calculate granular reducts of an inconsistent data set.

For the purpose of solving these problems, heuristic ideas are combined with reduction approach in our study so as to calculate reducts of an inconsistent formal decision context. Note that our reduction technique depends on consistency measure and it can be applicable to consistent formal decision contexts as well.

Our remainder work is described below. We briefly review the foundations of FCA in next section. And then, we discuss the issue of information granulation in formal contexts. Section 4 defines information entropy in formal contexts and conditional information entropy in formal decision contexts. Moreover, we use them to measure attribute significance. Section 5 investigates how to measure the consistency of a formal decision context which is helpful to the study of obtaining reducts from an inconsistent data set. In Section 6, heuristic ideas are combined with reduction technique to calculate reducts of an inconsistent data set.

2. Foundations of FCA

Definition 1. (Formal context, [39]). A formal context is described as (O, A, R), where R is an incidence relation between the object set O and the attribute set A. Moreover, we read $(o, b) \in R$ as "o has b" and $(o, b) \notin R$ as the opposite.

Given $M \subseteq O$ and $N \subseteq A$, we induce a pair of concept forming operators:

$$M' = \{b \in A | \forall o \in M, (o, b) \in R\},$$
$$N' = \{o \in O | \forall b \in N, (o, b) \in R\}.$$

Definition 2. (Formal concept, [39]). Let $\mathrm{K} = (O, A, R)$ be a formal context. For any $M \subseteq O$ and $N \subseteq A$, if $M' = N$ and $N' = M$, we say that (M, N) is a (formal) concept of K. M and N are referred to as the extent and the intent of (M, N), respectively.

If all concepts of K are equipped with the following subconcept-superconcept relation $(M_1, N_1) \prec (M_2, N_2) \Leftrightarrow M_1 \subseteq M_2$, they induce a concept lattice of K [39]. In the rest of this paper, we denote it by $\underline{L}(O, A, R)$. Moreover, the *meet* and *join* operators in $\underline{L}(O, A, R)$ are defined as

$(M_1, N_1) \wedge (M_2, N_2) = (M_1 \cap M_2, (N_1 \cup N_2)'')$ and $(M_1, N_1) \vee (M_2, N_2) = ((M_1 \cup M_2)'', N_1 \cap N_2)$.

Definition 3. (Subcontext, [40]). Let $\mathrm{K} = (O, A, R)$ be a formal context, $Q \subseteq A$ and write $R \cap (O \times Q)$ as R_Q. Then the formal context (O, Q, R_Q) is called a subcontext of K.

Similar to the case in K, we can define a pair of concept forming operators in (O, Q, R_Q). For $M \subseteq O$ and $N \subseteq Q$,

$$M^{'Q} = \{b \in Q | \forall o \in M, (o, b) \in R_Q\},$$
$$N^{'Q} = \{o \in O | \forall b \in N, (o, b) \in R_Q\}.$$

In fact, the above concept forming operators ${}^{'Q} : 2^O \to 2^Q$ and ${}^{'Q} : 2^Q \to 2^O$ are the restriction of the ones $' : 2^O \to 2^A$ and $' : 2^A \to 2^O$ on the subcontext (O, Q, R_Q), where 2^O, 2^Q and 2^A are the power sets of O, Q and A, respectively. Moreover, we call (M, N) a concept of (O, Q, R_Q) if $M^{'Q} = N$ and $N^{'Q} = M$, and use $\underline{L}(O, Q, R_Q)$ to represent the concept lattice of (O, Q, R_Q).

For brevity, $\{o\}^{'}_{Q}$ is represented as $o^{'}_{Q}$, and $\{b\}^{'}_{Q}$ as $b^{'}_{Q}$ for any $(o, b) \in O \times Q$.

Proposition 1. *[40]. Let (O, Q, R_Q) be a subcontext of $K = (O, A, R)$, $M, M_1, M_2 \subseteq O$ and $N, N_1, N_2 \subseteq Q$. We obtain the following statements:*

(1) $M_1 \subseteq M_2 \Rightarrow M_2^{'Q} \subseteq M_1^{'Q}, N_1 \subseteq N_2 \Rightarrow N_2^{'Q} \subseteq N_1^{'Q}$;

(2) $M \subseteq M^{'Q'Q}, N \subseteq N^{'Q'Q}$;

(3) $M^{'Q} = M^{'} \cap Q, N^{'Q} = N'$;

(4) $M^{''} \subseteq M^{'Q'Q}$;

(5) $(M^{'Q'Q}, M^{'Q}) \in \underline{L}(O, Q, R_Q)$.

It should be pointed out that the statements (1), (2) and (5) are similar to those in $\mathrm{K} = (O, A, R)$*. Moreover, (3) is trivial if we notice that* (O, Q, R_Q) *and* K *share the same object set, but the latter has more attributes than the former. Besides, (4) can be understood as "the minimal extent containing the granule M in* (O, Q, R_Q) *is greater than the one in* K*" since* K *has more extents than* (O, Q, R_Q)*. This is helpful to the study of information granulation because the information granules are known to become larger when attributes are removed from the original date set.*

Definition 4. (Formal decision context, [46]). A formal decision context (FDC) can be described as $\mathrm{S} = (O, A, R, D, J)$, where the contexts (O, A, R) and (O, D, J) satisfy $A \cap D = \emptyset$. A and D are called the conditional attribute set and the decision attribute set of S, respectively.

3. Information granulation in formal contexts

Note that Wu et al. [40] examined the granular structures of concept lattices and claimed that one can find information granules from a formal context such that they can determine the concept lattice of the formal context.

In this section, how to measure the degree of fineness (or coarseness) of the information granules is investigated in preparation for the subsequent study of information entropy and conditional information entropy.

First of all, the information granules are recalled in formal contexts. The concept forming operators $' : 2^O \to 2^A$ and $' : 2^A \to 2^O$ in $\mathrm{K} = (O, A, R)$ are sometimes rewritten as ${}^{\prime A}$ and ${}^{\prime A}$. In FCA, $(o^{\prime A \prime A}, o^{\prime A})$ $(o \in O)$ are called object concepts [18], and they can be used to induce any concept (M, N):

$$(M, N) = \vee_{o \in M}(o^{\prime A \prime A}, o^{\prime A}).$$

So, $\{(o^{\prime A \prime A}, o^{\prime A}) | o \in O\}$ can be a basis of $\underline{L}(O, A, R)$. Note that the extent $o^{\prime A \prime A}$ and the intent $o^{\prime A}$ of $(o^{\prime A \prime A}, o^{\prime A})$ are uniquely determined with each other [10]. Thus, $\{o^{\prime A \prime A} | o \in O\}$ is sufficient in terms of generating all formal concepts. Hereinafter, we call $\{o^{\prime A \prime A} | o \in O\}$, denoted by $\delta(O, A, R)$, the object-oriented information granules (simply information granules) of K. Obviously, the information granules $\delta(O, A, R)$ are a cover of O.

Definition 5. (Coarser or finer relation). For two formal contexts $\mathrm{K} = (O, A, R)$ and $\bar{\mathrm{K}} = (O, D, J)$, let $\delta(O, A, R)$ and $\delta(O, D, J)$ be their respective information granules. If $o^{\prime A \prime A} \subseteq o^{\prime D \prime D}$ for any $o \in O$, then $\delta(O, A, R)$ is said to be finer than $\delta(O, D, J)$ (or equivalently, $\delta(O, D, J)$ is coarser than $\delta(O, A, R)$). We represent this relationship by $\delta(O, A, R) \leq \delta(O, D, J)$. Furthermore, if $\delta(O, A, R) \leq \delta(O, D, J)$ and there exists $o_1 \in O$ such that $o_1^{\prime A \prime A} \subset o_1^{\prime D \prime D}$, we say that $\delta(O, A, R)$ is strictly finer than $\delta(O, D, J)$ (or equivalently, $\delta(O, D, J)$ is strictly coarser than $\delta(O, A, R)$). We denote this relationship by $\delta(O, A, R) < \delta(O, D, J)$.

Combining Definition 5 with Proposition 1, we obtain the following property.

Proposition 2. *Let* (O, Q, R_Q) *be a subcontext of* $\mathrm{K} = (O, A, R)$*. Then, information granules of* (O, Q, R_Q) *are coarser than those of* K*, i.e.* $\delta(O, A, R) \leq \delta(O, Q, R_Q)$*.*

Definition 6. (Information granulation). Let (O, Q, R_Q) be a subcontext of $\mathrm{K} = (O, A, R)$. Then, we define the information granulation (IG) of (O, Q, R_Q) as

$$\mathrm{IG}(Q) = \frac{1}{|O|} \sum_{o \in O} \frac{|o^{\prime Q \prime Q}|}{|O|}.$$

Example 1. Tab. I is a formal context $\mathrm{K} = (O, A, R)$, where $O = \{o_1, o_2, o_3, o_4, o_5\}$, $A = \{b_1, b_2, b_3, b_4, b_5, b_6, b_7\}$, and the signs "+" and "−" are respectively used to indicate "an object has an attribute" and "an object does not have an attribute".

O	b_1	b_2	b_3	b_4	b_5	b_6	b_7
o_1	+	−	+	−	−	−	+
o_2	−	+	−	−	−+	−	
o_3	−	+	+	−	+	−	−
o_4	−	−	−	+	+	+	−
o_5	+	−	−	−	−	−	−

Tab. I *A formal context* $\mathrm{K} = (O, A, R)$*.*

Take $Q = \{b_1, b_2, b_3, b_4\}$. We can obtain the information granules of the subcontext (O, Q, R_Q):

$$o_1^{\prime Q \prime Q} = \{o_1\}, \quad o_2^{\prime Q \prime Q} = \{o_2, o_3\}, \quad o_3^{\prime Q \prime Q} = \{o_3\}, \quad o_4^{\prime Q \prime Q} = \{o_4\}, \quad o_5^{\prime Q \prime Q} = \{o_1, o_5\}.$$

Furthermore, the relative size of $|o_i^{\prime Q^\prime Q}|$ to $|O|$ can be computed as follows:

$$\frac{\left|o_1^{\prime Q^\prime Q}\right|}{|O|} = 0.2, \quad \frac{\left|o_2^{\prime Q^\prime Q}\right|}{|O|} = 0.4, \quad \frac{\left|o_3^{\prime Q^\prime Q}\right|}{|O|} = 0.2, \quad \frac{\left|o_4^{\prime Q^\prime Q}\right|}{|O|} = 0.2, \quad \frac{\left|o_5^{\prime Q^\prime Q}\right|}{|O|} = 0.4.$$

Then, the average of the relative size of $\left|o_i^{\prime Q^\prime Q}\right|$ to $|O|$ is

$$\mathrm{IG}(Q) = \frac{1}{|O|} \sum_{o_i \in O} \frac{|o_i^{\prime Q^\prime Q}|}{|O|} = \frac{1}{5}(0.2 + 0.4 + 0.2 + 0.2 + 0.4) = \frac{7}{25}.$$

Based on Definition 6 and Example 1, we know that $\mathrm{IG}(Q)$ provides a useful approach to evaluate the degree of fineness (or coarseness) of information granules. The larger $\mathrm{IG}(Q)$ of a formal context is, the coarser the information granules of the formal context is. Moreover, it deserves to be mentioned that $\mathrm{IG}(Q)$ is in fact an average measure of the relative size of information granules.

Note that the value range of $\mathrm{IG}(Q)$ is between $1/|O|$ and 1. Besides, according to Definition 6 and the fourth item of Proposition 1, the following statement is true.

Proposition 3. *Let* (O, Q, R_Q) *be a subcontext of* $\mathrm{K} = (O, A, R)$. *Then* $IG(A) \leq IG(Q)$.

Thus, when some attributes are removed from $\mathrm{K} = (O, A, R)$, the information granules $\delta(O, A, R)$ will become coarser, and the difference $\mathrm{IG}(Q) - \mathrm{IG}(A)$ reflects the change of the degree of coarseness of $\delta(O, A, R)$.

Example 2. Continued with Example 1. The information granules of $\mathrm{K} = (O, A, R)$ are as follows:

$$o_1^{\prime A^\prime A} = \{o_1\}, \quad o_2^{\prime A^\prime A} = \{o_2\}, \quad o_3^{\prime A^\prime A} = \{o_3\}, \quad o_4^{\prime A^\prime A} = \{o_4\}, \quad o_5^{\prime A^\prime A} = \{o_1, o_5\}.$$

By Definition 6, we obtain

$$\mathrm{IG}(A) = \frac{1}{|O|} \sum_{o_i \in O} \frac{|o_i^{\prime A^\prime A}|}{|O|} = \frac{1}{5}(0.2 + 0.2 + 0.2 + 0.2 + 0.4) = \frac{6}{25}.$$

Hence, when the attributes b_5, b_6 and b_7 are removed from K, the information granules of K will become coarser, and the difference $\mathrm{IG}(Q) - \mathrm{IG}(A) = \frac{1}{25}$ indicates the change of the degree of coarseness of $\delta(O, A, R)$.

4. Information entropy and attribute significance

As is well known in rough set theory, entropy theory was used to deal with the problem of measuring attribute significance [19, 22, 36]. In this section, we employ entropy theory to solve the similar problem in FCA.

In what follows, we first propose the notions of information entropy in formal contexts and conditional information entropy in formal decision contexts based on information granules. Furthermore, we use them for the evaluation of the significance of attributes.

Definition 7. (Information entropy). Let (O, Q, R_Q) be a subcontext of $\mathrm{K} = (O, A, R)$. Then, we define the information entropy (IE) of (O, Q, R_Q) as

$$\mathrm{IE}(Q) = \sum_{o \in O} \frac{1}{|O|} \left(1 - \frac{|o'^{Q'Q}|}{|O|} \right).$$

In fact, the information entropy $\mathrm{IE}(Q)$ is a measure of the uncertainty of information provided by (O, Q, R_Q) which is defined based on information granules. More explanations are as follows:

(a) $o^{'Q'Q}$ $(o \in O)$ can be viewed as all the events of a sampling test and they have the same probability $\frac{1}{|O|}$;

(b) the information entropy $\mathrm{IE}(Q)$ reflects the uncertainty of a sampling test, and it can be used to measure the quantity of information provided by (O, Q, R_Q);

(c) when $o^{'Q'Q}$ $(o \in O)$ are all singleton sets, the information entropy $\mathrm{IE}(Q)$ reaches the maximum value. In this case, all the events $o^{'Q'Q}$ $(o \in O)$ are the most unstable and the sampling test gets the biggest uncertainty.

(d) when $o^{'Q'Q}$ $(o \in O)$ are O, the information entropy $\mathrm{IE}(Q)$ reaches the minimum value. In this case, $o^{'Q'Q}$ $(o \in O)$ are all certain events and the sampling test has no uncertainty.

Proposition 4. *Let* (O, Q, R_Q) *be a subcontext of* $\mathrm{K} = (O, A, R)$. *Then* $\mathrm{IG}(Q) + \mathrm{IE}(Q) = 1$.

Proof. It can be obtained directly by Definitions 6 and 7. □

Based on Proposition 4 and Definitions 5 and 7, the coarser the information granules of a formal context are, the less the uncertainty of information provided by the formal context is.

Proposition 5. *Let* (O, Q, R_Q) *be a subcontext of* $\mathrm{K} = (O, A, R)$. *Then* $\mathrm{IE}(Q) \leq \mathrm{IE}(A)$.

Proof. It can be obtained directly by Definition 7 and (4) of Proposition 1. □

Proposition 5 says that the uncertainty of information provided by a formal context will decrease when some attributes are removed, since the information granules become coarser.

Definition 8. (Conditional information entropy). Let $\mathrm{S} = (O, A, R, D, J)$ be a FDC. Then, we define the conditional information entropy (CIE) of (O, D, J) to (O, A, R) as

$$\mathrm{CIE}(D|A) = \sum_{o \in O} \frac{1}{|O|} \frac{|o'^{A'A}| - |o'^{A'A} \cap o'^{D'D}|}{|O|}.$$

In fact, the conditional information entropy of (O, D, J) to (O, A, R) indicates the information entropy of $(O, A \cup D, R \cup J)$ under the condition that the information provided by (O, A, R) in the form of information granules has been known. In other words, the following statement is true.

Proposition 6. *Let* $S = (O, A, R, D, J)$ *be a FDC. We have* $\mathrm{CIE}(D|A) = \mathrm{IE}(A \cup D) - \mathrm{IE}(A)$*, where* $\mathrm{IE}(A \cup D)$ *is the information entropy of* $(O, A \cup D, R \cup J)$*.*

Proof. From Definitions 7 and 8, we have

$$\begin{aligned}
\mathrm{IE}(A \cup D) - \mathrm{IE}(A) &= \sum_{o \in O} \frac{1}{|O|}\left(1 - \frac{|o'^{(A\cup D)'(A \cup D)}|}{|O|}\right) - \sum_{o \in O} \frac{1}{|O|}\left(1 - \frac{|o'^{A'A}|}{|O|}\right) \\
&= \sum_{o \in O} \frac{1}{|O|}\left(1 - \frac{|o'^{A'A} \cap o'^{D'D}|}{|O|}\right) - \sum_{o \in O} \frac{1}{|O|}\left(1 - \frac{|o'^{A'A}|}{|O|}\right) \\
&= \sum_{o \in O} \frac{1}{|O|} \frac{|o'^{A'A}| - |o'^{A'A} \cap o'^{D'D}|}{|O|} \\
&= \mathrm{CIE}(D|A).
\end{aligned}$$

□

Proposition 7. *Let* $\mathrm{S} = (O, A, R, D, J)$ *be a FDC. We have* $\mathrm{IE}(D) \geq \mathrm{CIE}(D|A)$*.*

Proof. From Definitions 7 and 8, we have

$$\begin{aligned}
\mathrm{IE}(D) - \mathrm{CIE}(D|A) &= \sum_{o \in O} \frac{1}{|O|}\left(1 - \frac{|o'^{D'D}|}{|O|}\right) - \sum_{o \in O} \frac{1}{|O|} \frac{|o'^{A'A}| - |o'^{A'A} \cap o'^{D'D}|}{|O|} \\
&= \sum_{o \in O} \frac{1}{|O|}\left(1 - \frac{|o'^{D'D}| + |o'^{A'A}| - |o'^{A'A} \cap o'^{D'D}|}{|O|}\right) \\
&= \sum_{o \in O} \frac{1}{|O|}\left(1 - \frac{|o'_{A'A} \cup o'^{D'D}|}{|O|}\right) \\
&\geq 0.
\end{aligned}$$

Therefore, $\mathrm{IE}(D) \geq \mathrm{CIE}(D|A)$. □

In what follows, we use $\mathrm{IE}(Q)$ for the evaluation of the significance of attributes of a formal context.

Definition 9. (Inner significance). Let $\mathrm{K} = (O, A, R)$ be a formal context. Then, we define the significance of b in A as

$$\mathrm{SIG}(A|b) = \mathrm{IE}(A) - \mathrm{IE}(A \backslash \{b\}).$$

It can be observed that the significance of b in A is evaluated by the magnitude that the information entropy of $\mathrm{K} = (O, A, R)$ changes when the attribute b is removed from A. This kind of attribute significance provides a quantitative analysis of the contribution that each attribute of A makes to $\mathrm{K} = (O, A, R)$.

Definition 10. (Outer significance). Let $\mathrm{K} = (O, A, R)$ be a formal context and $Q \subset A$. We define the significance of $b \in A \backslash Q$ with respect to Q as

$$\mathrm{SIG}(b|Q) = \mathrm{IE}(Q \cup \{b\}) - \mathrm{IE}(Q).$$

From Definition 10, we know that $\mathrm{SIG}(b|Q)$ is measured by the magnitude that the information entropy of (O, Q, R_Q) changes when the attribute b is added into Q. This new kind of attribute significance provides a quantitative analysis of the contribution that each attribute in $A\backslash Q$ will make to (O, Q, R_Q) if it is added into Q.

Similarly, we can use conditional information entropy for the evaluation of the significance of attributes of a FDC. We leave this issue to be discussed in Section 6.

Example 3. Tab. II is a formal context $\bar{\mathrm{K}} = (O, D, J)$, where $O = \{o_1, o_2, o_3, o_4, o_5\}$ and $D = \{c_1, c_2, c_3, c_4\}$.

O	c_1	c_2	c_3	c_4
o_1	+	−	+	−
o_2	+	+	−	−
o_3	+	+	−	−
o_4	−	+	−	
o_5	−	−	−	+

Tab. II *A formal context* $\bar{\mathrm{K}} = (O, D, J)$.

The information granules of $\bar{\mathrm{K}}$ are as follows:
$o_1^{\prime D^\prime D} = \{o_1\}, \quad o_2^{\prime D^\prime D} = \{o_2, o_3\}, \quad o_3^{\prime D^\prime D} = \{o_2, o_3\}, \quad o_4^{\prime D^\prime D} = \{o_4\}, \quad o_5^{\prime D^\prime D} = \{o_5\}.$
According to Definition 7, the information entropy of $\bar{\mathrm{K}}$ is

$$\mathrm{IE}(D) = \sum_{o \in O} \frac{1}{|O|} \left(1 - \frac{|o^{\prime D^\prime D}|}{|O|}\right) = \frac{1}{5}\,(0.8 + 0.6 + 0.6 + 0.8 + 0.8) = \frac{18}{25}.$$

Furthermore, let us also concern $\mathrm{K} = (O, A, R)$ in Tab. I. Then, we can obtain

$$\mathrm{IE}(A) = \frac{19}{25} \text{ and } \mathrm{IE}(A \cup D) = \frac{4}{5}.$$

By Definition 8, we know that the conditional information entropy of $\bar{\mathrm{K}} = (O, D, J)$ to $\mathrm{K} = (O, A, R)$ is

$$\mathrm{CIE}(D|A) = \sum_{o \in O} \frac{1}{|O|} \frac{|o^{\prime A^\prime A}| - |o^{\prime A^\prime A} \cap o^{\prime D^\prime D}|}{|O|} = \frac{1}{25}.$$

Thus, $\mathrm{CIE}(D|A) = \mathrm{IE}(A \cup D) - \mathrm{IE}(A)$ and $\mathrm{IE}(D) \geq \mathrm{CIE}(D|A)$.

Besides, the significance of attributes of $\bar{\mathrm{K}} = (O, D, J)$is shown below one by one:

$$\mathrm{SIG}(D|c_1) = \mathrm{IE}(D) - \mathrm{IE}(D\backslash\{c_1\}) = \frac{18}{25} - \frac{15}{25} = \frac{3}{25},$$
$$\mathrm{SIG}(D|c_2) = \mathrm{IE}(D) - \mathrm{IE}(D\backslash\{c_2\}) = \frac{18}{25} - \frac{15}{25} = \frac{3}{25},$$
$$\mathrm{SIG}(D|c_3) = \mathrm{IE}(D) - \mathrm{IE}(D\backslash\{c_3\}) = \frac{18}{25} - \frac{14}{25} = \frac{4}{25},$$
$$\mathrm{SIG}(D|c_4) = \mathrm{IE}(D) - \mathrm{IE}(D\backslash\{c_4\}) = \frac{18}{25} - \frac{14}{25} = \frac{4}{25}.$$

5. Consistency measure of FDCs

Definition 11. (Consistency, [40]). Let $\mathrm{S}=(O,A,R,D,J)$ be a FDC. If $o^{'A'A}\subseteq o^{'D'D}$ holds for any $o\in O$, we say that S is consistent; otherwise, it is inconsistent.

Consistent FDCs are in fact a special type of FDCs since any consistent FDC $\mathrm{S}=(O,A,R,D,J)$ satisfies that the information granules of (O,A,R) are finer than those of (O,D,J), i.e. $\delta(O,A,R)\leq\delta(O,D,J)$. The particularity of consistent FDCs can guarantee that for every $o\in O$, we can induce a granular rule $o^{'A}\to o^{'D}$, which, in turn, makes the consistent FDCs quite important [40]. However, the possibility of an inconsistent FDC appearing is more than that of a consistent one in the real world. Then it is natural to ask how to evaluate the degree of consistency of an inconsistent FDC? This problem is very important because different inconsistent FDCs have difference in the degree of consistency and it is inappropriate to just view them as the same. In what follows, we try to address this problem.

For a FDC $\mathrm{S}=(O,A,R,D,J)$ and $Q\subseteq A$, we denote

$$\mathrm{POS}_Q(D)=\{o\in O|o^{'Q'Q}\subseteq o^{'D'D}\}.$$

That is, $\mathrm{POS}_Q(D)$ is constituted by the objects whose induced information granules under the subcontext (O,Q,R_Q) are finer than their corresponding information granules under (O,D,J). Note that S is a consistent FDC if and only if $\mathrm{POS}_A(D)=O$ holds. So, we use the ratio of $\mathrm{POS}_Q(D)$ to $|O|$ in order to evaluate the degree of consistency of a FDC.

Definition 12. (Consistency degree). Let $\mathrm{S}=(O,A,R,D,J)$ be a FDC and $Q\subseteq A$. The degree of consistency of the subcontext (O,Q,R_Q,D,J) of S is measured by

$$\tau_Q(D)=\frac{|\mathrm{POS}_Q(D)|}{|O|}.$$

Based on Proposition 1 and Definitions 11 and 12, the following statements are true.

Proposition 8. *Let* $\mathrm{S}=(O,A,R,D,J)$ *be a FDC. Then* S *is consistent iff* $\tau_Q(D)=1$.

Proposition 9. *Let* $\mathrm{S}=(O,A,R,D,J)$ *be a FDC and* $Q\subseteq A$. *Then* $\tau_Q(D)\leq\tau_A(D)$.

Proposition 9 says that the degree of consistency of a FDC will decrease when some of its conditional attributes are removed.

Finally, we use the degree of consistency to represent the reduction of consistent FDCs, which will make the discussion of this issue in inconsistent FDCs more natural.

Definition 13. (Granular consistent set, [40]). Let $\mathrm{S}=(O,A,R,D,J)$ be a consistent FDC and $Q\subseteq A$. If $o^{'Q'Q}\subseteq o^{'D'D}$ holds for any $o\in O$, then Q is called a granular consistent set (GCS) of S. If Q is a GCS of S and any proper subset of Q is not a GCS of S, then Q is called a granular reduct of S.

Proposition 10. *Let* $\mathrm{S} = (O, A, R, D, J)$ *be a consistent FDC and* $Q \subseteq A$. Q *is a* GCS *of* S *iff* $\tau_Q(D) = \tau_A(D)$.

Proof. It can be obtained directly by Definitions 12 and 13. □

Proposition 11. *Let* $\mathrm{S} = (O, A, R, D, J)$ *be a consistent FDC and* $Q \subseteq A$. Q *is a granular reduct of* S *iff* $\tau_Q(D) = \tau_A(D)$ *and* $\forall b \in Q$, $\tau_{Q\setminus\{b\}}(D) < \tau_Q(D)$.

Proof. It is trivial. □

Proposition 11 gives a new version for simplification of a consistent FDC from the perspective of the degree of consistency. Specifically, a granular reduct of a consistent FDC is a minimal GCS preserving the degree of consistency.

6. Attribute reduction in inconsistent FDCs

In the previous section, we have shown that attribute reduction of a consistent FDC can be represented by the degree of consistency. Now, we discuss the notion of attribute reduction in inconsistent FDCs.

Definition 14. (Granular reduct). Let $\mathrm{S} = (O, A, R, D, J)$ be an inconsistent FDC and $Q \subseteq A$. If $\tau_Q(D) = \tau_A(D)$, then Q is called a generalized GCS of S. If Q is a generalized GCS of S and any proper subset of Q is not a generalized GCS of S, then Q is called a granular reduct of S. We call $\cap Q_t$ the core of S and denote it by Core(S).

A granular reduct of an inconsistent FDC is a minimal generalized GCS preserving the degree of consistency, which is in accordance with the one in consistent FDC. Moreover, by Proposition 11 and Definition 14, the simplification of inconsistent FDCs can be viewed as a natural generalization of the one in consistent FDCs. So, any results obtained below on the simplification of inconsistent FDCs are automatically suitable for that of consistent FDCs.

In Section 4, we have used information entropy for the evaluation of the significance of attributes of $\mathrm{K} = (O, A, R)$. In what follows, we propose the notion of limitary (conditional) information entropy in inconsistent FDCs for the evaluation of attribute significance.

Definition 15. (Limitary information entropy). Let $\mathrm{S} = (O, A, R, D, J)$ be an inconsistent FDC and $Q \subseteq A$. Then, we define the limitary information entropy (LIE) of (O, Q, R_Q) as

$$\mathrm{LIE}(Q) = \sum_{o \in \mathrm{POS}_A(D)} \frac{1}{|O|} \left(1 - \frac{|o'^{Q'Q}|}{|O|} \right)$$

and the limitary conditional information entropy (LCIE) of (O, D, J) to (O, Q, R_Q) as

$$\mathrm{LCIE}(D|Q) = \sum_{o \in \mathrm{POS}_A(D)} \frac{1}{|O|} \frac{|o'^{Q'Q}| - |o'^{Q'Q} \cap o'^{D'D}|}{|O|}.$$

Proposition 12. *Let* $\mathrm{S} = (O, A, R, D, J)$ *be an inconsistent FDC and* $P \subseteq Q \subseteq A$. *Then* $\mathrm{LCIE}(D|Q) \leq \mathrm{LCIE}(D|P)$.

Proof. The proof follows immediately from Definition 15. □

Definition 16. (Inner attribute significance). Let $\mathrm{S} = (O, A, R, D, J)$ be an inconsistent FDC and $b \in Q \subseteq A$. Then, the significance of b in Q is defined as

$$\mathrm{SIG}(Q|b) = \mathrm{LCIE}(D|Q\backslash\{b\}) - \mathrm{LCIE}(D|Q).$$

Based on Definition 16, the significance of b in Q is measured by the magnitude that the limitary conditional information entropy of (O, D, J) to (O, Q, R_Q) changes when the attribute b is removed from Q. This kind of attribute significance provides a quantitative analysis of the contribution that each attribute of Q makes to the inconsistent FDC (O, Q, R_Q, D, J).

Definition 17. (Outer attribute significance). Let $\mathrm{S} = (O, A, R, D, J)$ be an inconsistent FDC and $Q \subset A$. The significance of $b \in A\backslash Q$ with respect to Q is defined as

$$\mathrm{SIG}(b|Q) = \mathrm{LCIE}(D|Q) - \mathrm{LCIE}(D|Q \cup \{b\}).$$

It can be known from Definition 17 that the significance of $b \in A\backslash Q$ with respect to Q can be evaluated by the value that the limitary conditional information entropy of (O, D, J) to (O, Q, R_Q) changes when the attribute b is added into Q. This new kind of attribute significance provides a quantitative analysis of the contribution that each attribute in $A\backslash Q$ will make to (O, Q, R_Q, D, J) if it is added into Q.

Furthermore, we use the above two kinds of attribute significance and limitary conditional information entropy to discuss the equivalent conditions of generalized GCS, granular reduct and core in inconsistent FDCs.

Proposition 13. *Let* $\mathrm{S} = (O, A, R, D, J)$ *be an inconsistent FDC and* $Q \subseteq A$. *Then* Q *is a generalized GCS of* S *iff* $\mathrm{LCIE}(D|Q) = 0$.

Proof. Necessity. If Q is a generalized GCS of S, we obtain $\tau_Q(D) = \tau_A(D)$ from Definition 14, which implies $\mathrm{POS}_Q(D) = \mathrm{POS}_A(D)$. Thus, for any $o \in \mathrm{POS}_A(D)$, it follows $o \in \mathrm{POS}_Q(D)$, i.e., $o'^{Q'Q} \subseteq o'^{D'D}$. This leads to

$$\mathrm{LCIE}(D|Q) = \sum_{o \in \mathrm{POS}_A(D)} \frac{1}{|O|} \frac{|o'^{Q'Q}| - |o'^{Q'Q} \cap o'^{D'D}|}{|O|} = 0.$$

Sufficiency. If $\mathrm{LCIE}(D|Q) = 0$, then we have $o'^{Q'Q} \subseteq o'^{D'D}$ for all $o \in \mathrm{POS}_A(D)$. Thus, $o \in \mathrm{POS}_Q(D)$ holds for all $o \in \mathrm{POS}_A(D)$, which yields $\mathrm{POS}_A(D) \subseteq \mathrm{POS}_Q(D)$. Furthermore, noting that $\mathrm{POS}_Q(D) \subseteq \mathrm{POS}_A(D)$ is satisfied due to $Q \subseteq A$, we conclude $\mathrm{POS}_Q(D) = \mathrm{POS}_A(D)$. Therefore, $\tau_Q(D) = \tau_A(D)$. By Definition 14, Q is a generalized GCS of S. □

Theorem 14. *Let* $\mathrm{S} = (O, A, R, D, J)$ *be an inconsistent FDC and* $Q \subseteq A$. R *is a granular reduct of* S *iff* $LCIE(D|Q) = 0$ *and* $\forall b \in Q$, $LCIE(D|Q\backslash\{b\}) > 0$.

Proof. Necessity. If Q is a granular reduct of S, then Q is a generalized GCS of S and any proper subset of Q is not a generalized GCS of S. By Proposition 13, it follows $\mathrm{LCIE}(D|Q) = 0$. Furthermore, $\mathrm{LCIE}(D|Q\backslash\{b\}) > 0$ is also true for any $b \in Q$. Otherwise, there is $b_0 \in Q$ satisfying $\mathrm{LCIE}(D|Q\backslash\{b_0\}) = 0$, then it can be known from Proposition 13 that $Q\backslash\{b_0\}$ is a generalized GCS of S. This is in contradiction with the assumption that any proper subset of Q is not a generalized GCS of S.

Sufficiency. Since $\mathrm{LCIE}(D|Q) = 0$, we have that Q is a generalized GCS of S. Note that for any $P \subset Q$, there must exist $b_0 \in Q\backslash P$ such that $P \subseteq Q\backslash\{b_0\}$. Based on Proposition 12 and the assumption, we obtain $\mathrm{LCIE}(D|P) \geq \mathrm{LCIE}(D|Q\backslash\{b_0\}) > 0$. Based on Proposition 13, P is not a generalized GCS of S. Consequently, Q is a granular reduct of S. □

Theorem 15. *Let* $\mathrm{S} = (O, A, R, D, J)$ *be an inconsistent FDC. Then* $Core(S) = \{b \in A|SIG(A|b) > 0\}$.

Proof. Suppose all granular reducts of S are $\{Q_t|t \in T\}$. According to Definition 14, it follows $\mathrm{Core}(S) = \bigcap_{t\in T} Q_t$.

Firstly, we prove $\mathrm{Core}(S) \subseteq \{b \in A|\mathrm{SIG}(A|b) > 0\}$. For any $c \in \mathrm{Core}(S)$, we have $c \in Q_t$ for all $t \in T$ due to $\mathrm{Core}(S) = \cap_{t\in T} Q_t$. If $\mathrm{SIG}(A|c) = 0$, i.e. $\mathrm{LCIE}(D|A\backslash\{c\}) = \mathrm{LCIE}(D|A)$, then $\mathrm{LCIE}(D|A\backslash\{c\}) = 0$ since $\mathrm{LCIE}(D|A) = 0$. By Proposition 13, $A\backslash\{c\}$ is a generalized GCS of S. Note that every generalized GCS of S has at least one granular reduct of S. So, we can find $Q_{t_0} \subseteq A\backslash\{c\}$ which is a granular reduct of S. However, $c \in Q_{t_0} \subseteq A\backslash\{c\}$, a contradiction. Therefore, $c \in \{b \in A|\mathrm{SIG}(A|b) > 0\}$.

Secondly, we prove $\{b \in A|\mathrm{SIG}(A|b) > 0\} \subseteq \mathrm{Core}(S)$. For any $c \in \{b \in A|\mathrm{SIG}(A|b) > 0\}$, we have $\mathrm{SIG}(A|c) > 0$, i.e., $\mathrm{LCIE}(D|A\backslash\{c\}) > 0$. If $c \notin \mathrm{Core}(S)$, then there exists a granular reduct Q_{t_0} of S such that $c \notin Q_{t_0}$, yielding $Q_{t_0} \subseteq A\backslash\{c\}$. From Proposition 12, we obtain $\mathrm{LCIE}(D|Q_{t_0}) \geq \mathrm{LCIE}(D|A\backslash\{c\}) > 0$, which is in contradiction with the known result that Q_{t_0} is a granular reduct of S. □

On the basis of the above discussion, we put forward a heuristic reduction procedure for the acquisition of granular reducts from an inconsistent FDC. The rationale of the forthcoming procedure is described below: For an inconsistent FDC $\mathrm{S} = (O, A, R, D, J)$, it is known from Definition 14 that Core(S) of S is included in every granular reduct of S. Therefore, in order to enhance the efficiency of computing a granular reduct, we start with Core(S) to generate a granular reduct. Furthermore, if $\mathrm{LCIE}(D|\mathrm{Core}(A)) = 0$, then by Proposition 13 Core(S) is a generalized GCS of S; otherwise, we choose such an attribute c from A\Core(S) that can make the greatest contribution to $(O, \mathrm{Core(S)}, R_{\mathrm{Core(S)}}, D, J)$ with respect to the remainder in A\Core(S), i.e.,

$$\mathrm{SIG}(c|\mathrm{Core(S)}) = \max_{b\in A\backslash\mathrm{Core(S)}}\{\mathrm{SIG}(b|\mathrm{Core(S)})\},$$

and add the attribute c into Core(S) since such an addition strategy is the quickest way to drop $\mathrm{LCIE}(D|\mathrm{Core(S)})$ according to Definition 17. This process is repeated until $\mathrm{LCIE}(D|\mathrm{Core(S)}) = 0$. That is, Core(S) is a generalized GCS of S. Furthermore, by Theorem 14, any attribute d with $\mathrm{SIG}(\mathrm{Core(S)}|d) = 0$ should be removed from Core(S) one by one and a granular reduct can be obtained. The above process is represented by the following procedure.

Algorithm 1 Compute a reduct of an inconsistens FDC $\mathrm{S} = (O, A, R, D, J)$.

Ititialize $\mathrm{Core}(S) = \emptyset$.
repeat
  if $\mathrm{SIG}(A|b) > 0$ **then**
    $\mathrm{Core(S)} \leftarrow \mathrm{Core(S)} \cup \{b\}$.
  end if
until all elements of A have been checked
repeat
  Choose such an attribute c from $A \backslash \mathrm{Core(S)}$ that satisfies

$$\mathrm{SIG}(c|\mathrm{Core(S)}) = \max_{b \in A \backslash \mathrm{Core(S)}} \{\mathrm{SIG}(b|\mathrm{Core(S)})\},$$

  and add c into Core(S).
until $\mathrm{LCIE}(D|\mathrm{Core(S)}) = 0$
repeat
  if $\mathrm{SIG}(\mathrm{Core(S)}|d) = 0$ **then**
    $\mathrm{Core(S)} \leftarrow \mathrm{Core(S)} \backslash \{d\}$
  end if
until all elements of Core(S) have been checked
Output Core(S)

The time complexity of Algorithm 1 is $O(|O|^2(|A|^3 + |D|))$.

Example 4. Let (O, A, R) and (O, D, J) be the formal contexts in Tabs. I and II, respectively. Then $\mathrm{S} = (O, A, R, D, J)$ is a FDC. Since

$$o_5'^{A'A} = \{o_1, o_5\} \not\subseteq \{o_5\} = o_5'^{D'D},$$

it follows from Definitions 11 and 12 that S is an inconsistent FDC and the degree of consistency is 0.8.

In what follows, we use Algorithm 1 to perform the simplification of S. The significance of each attribute of A is as follows:

$$\mathrm{SIG}(A|b_2) = \frac{1}{25}, \quad \mathrm{SIG}(A|b_i) = 0 (i = 1, 3, 4, 5, 6, 7).$$

Thus, according to Theorem 15, the core of S is $\mathrm{Core(S)} = \{b_2\}$ and we can further calculate

$$\mathrm{LCIE}(D|\mathrm{Core}(A)) = \frac{8}{25} \neq 0.$$

Based on Theorem 14, more attributes need to be added into Core(S) to generate a granular reduct of S. By Definition 17, the significance of each attribute in $\{b_1, b_3, b_4, b_5, b_6, b_7\}$ with respect to Core(S) is as follows:

$$\mathrm{SIG}(b_1|\mathrm{Core(S)}) = \frac{3}{25}, \quad \mathrm{SIG}(b_3|\mathrm{Core(S)}) = \frac{3}{25}, \quad \mathrm{SIG}(b_4|\mathrm{Core(S)}) = \frac{4}{25},$$

$$\mathrm{SIG}(b_5|\mathrm{Core(S)}) = \frac{3}{25}, \quad \mathrm{SIG}(b_6|\mathrm{Core(S)}) = \frac{3}{25}, \quad \mathrm{SIG}(b_7|\mathrm{Core(S)}) = \frac{4}{25}.$$

We choose b_4 and add it into Core(S), i.e., $P = \mathrm{Core(S)} \cup \{b_4\} = \{b_2, b_4\}$. Since

$$\mathrm{LCIE}(D|P) = \mathrm{LCIE}(D|\{b_2, b_4\}) = \frac{4}{25} \neq 0,$$

we still need to choose some attributes from $\{b_1, b_3, b_5, b_6, b_7\}$ and add them into P for the computation of a granular reduct. According to Definition 17, we have

$$\mathrm{SIG}(b_1|P) = \frac{3}{25}, \quad \mathrm{SIG}(b_3|P) = \frac{3}{25},$$

$$\mathrm{SIG}(b_7|P) = \frac{4}{25}, \quad \mathrm{SIG}(b_i|P) = 0(i = 5, 6).$$

Therefore, we choose b_7 and add it into P, i.e., $Q = P \cup \{b_7\} = \{b_2, b_4, b_7\}$. Then, we find that $Q = \{b_2, b_4, b_7\}$ is a granular reduct of S since $\mathrm{LCIE}(D|Q) = 0$ and $\mathrm{LCIE}(D|Q\backslash\{b\}) > 0$ for any $b \in Q$.

It should be pointed out that the degree of consistency of the reduced inconsistent FDC (O, Q, R_Q, D, J) is still 0.8. In other words, although the attributes b_1, b_3, b_5 and b_6 have been removed from A, the degree of consistency keeps invariant.

7. A brief summary

This paper has proposed the notions of information entropy in formal contexts and conditional information entropy in FDCs with the help of information granules. Also we have used them for the evaluation of the attribute significance. Moreover, an investigation has been made on the consistency measure of a FDC, and a heuristic reduction technique has been developed for inconsistent FDCs. The proposed reduction method has polynomial-time complexity and can be applicable to consistent FDCs as well.

In fact, how can we evaluate the uncertainty of knowledge derived from a formal context or a FDC is a key problem in FCA. However, many challenging problems also appear: (1) Where is the uncertainty from? (2) Is the uncertainty from the data itself [1,4,16,32–34] or the instability of concepts [48]? (3) How can we measure the uncertainty objectively? Information entropy and conditional information entropy may provide some alternative ways to measure the uncertainty in formal concept analysis.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Nos. 61305057, 61562050 and 61573173) and Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province (No. OBDMA201502). The third author, Dr. Sérgio M. Dias, would like to thank the partial support of the Federal Service of Data Processing – www.serpro.gov.br.

References

[1] ASWANI KUMAR Ch. Fuzzy clustering-based formal concept analysis for association rules mining. *Applied Artificial Intelligence*. 2012, 26(3), pp. 274–301, doi: 10.1080/08839514.2012.648457.

[2] ASWANI KUMAR Ch., DIAS S.M., VIEIRA N.J. Knowledge reduction in formal contexts using non-negative matrix factorization. *Mathematics and Computers in Simulation*. 2015, 109, pp. 46–63, doi: 10.1016/j.matcom.2014.08.004.

[3] ASWANI KUMAR Ch., SRINIVAS S. Mining associations in health care data using formal concept analysis and singular value decomposition. *Journal of Biological Systems*. 2010, 18(4), pp. 787–807, doi: 10.1142/S0218339010003512.

[4] ASWANI KUMAR Ch., SRINIVAS S. Concept lattice reduction using fuzzy K-means clustering. *Expert Systems with Applications*. 2010, 37(3), pp. 2696–2704, doi: 10.1016/j.eswa.2009.09.026.

[5] CARPINETO C., ROMANO G. Exploiting the potential of concept lattices for information retrieval with CREDO. *Journal of Universal Computer Science*. 2004, 10(8), pp. 985–1013, doi: 10.3217/jucs-010-08-0985.

[6] DIAS S.M., VIEIRA N.J. Concept lattices reduction: definition, analysis and classification. *Expert Systems with Applications*. 2015, 42(20), pp. 7084–7097, doi: 10.1016/j.eswa.2015.04.044.

[7] DIAS S.M., VIEIRA N.J. Applying the JBOS reduction method for relevant knowledge extraction. *Expert Systems with Applications*. 2013, 40(5), pp. 1880–1887, doi: 10.1016/j.eswa.2012.10.010.

[8] DIAS S.M., ZÁRATE L.E., VIEIRA N.J. Extracting reducible knowledge from ANN with JBOS and FCANN approaches. *Expert Systems with Applications*. 2013, 40(8), pp. 3087–3095, doi: 10.1016/j.eswa.2012.12.024.

[9] ELLOUMI S., JAAM J., HASNAH A., JAOUA A., NAFKHA I. A multi-level conceptual data reduction approach based on the Lukasiewicz implication. *Information Sciences*. 2004, 163(4), pp. 253–262, doi: 10.1016/j.ins.2003.06.013.

[10] GANTER B., WILLE R. *Formal Concept Analysis: Mathematical Foundations*. Berlin: Springer, 1999.

[11] KUZNETSOV S.O. Machine learning on the basis of formal concept analysis. *Automation and Remote Control*. 2001, 62(10), pp. 1543–1564, doi: 10.1023/A:1012435612567.

[12] LI J.L., HE Z.Y., ZHU X.L. An entropy-based weighted concept lattice for merging multi-source geo-ontologies. *Entropy*. 2013, 15(6), pp. 2303–2318, doi: 10.3390/e15062303.

[13] LI J., HUANG C., QI J., QIAN Y., LIU W. Three-way cognitive concept learning via multi-granularity. *Information Sciences*. 2017, 378, pp. 244–263, doi: 10.1016/j.ins.2016.04.051.

[14] LI J., MEI C., ASWANI KUMAR Ch., ZHANG X. On rule acquisition in decision formal contexts. *International Journal of Machine Learning and Cybernetics*. 2013, 4(6), pp. 721–731, doi: 10.1007/s13042-013-0150-z.

[15] LI J., MEI C., LV Y. Knowledge reduction in decision formal contexts. *Knowledge-Based Systems*. 2011, 24(5), pp. 709–715, doi: 10.1016/j.knosys.2011.02.011.

[16] LI J., MEI C., LV Y. Incomplete decision contexts: Approximate concept construction, rule acquisition and knowledge reduction. *International Journal of Approximate Reasoning*. 2013, 54(1), pp. 149–165, doi: 10.1016/j.ijar.2012.07.005.

[17] LI J., MEI C., XU W., QIAN Y. Concept learning via granular computing: a cognitive viewpoint. *Information Sciences*. 2015, 298, pp. 447–467, doi: 10.1016/j.ins.2014.12.010.

[18] LI J., MEI C., WANG J., ZHANG X. Rule-preserved object compression in formal decision contexts using concept lattices. *Knowledge-Based Systems*. 2014, 71, pp. 435–445, doi: 10.1016/j.knosys.2014.08.020.

[19] LIANG J.Y., QIAN Y.H. Information granules and entropy theory in information systems. *Science in China Series F: Information Sciences*. 2008, 51(10), pp. 1427–1444, doi: 10.1007/s11432-008-0113-2.

[20] MA J.M., ZHANG W.X., LEUNG Y., SONG X.X. Granular computing and dual Galois connection. *Information Sciences*. 2007, 177, pp. 5365–45377, doi: 10.1016/j.ins.2007.07.008.

[21] MI J.S., LEUNG Y., WU W.Z. Approaches to attribute reduction in concept lattices induced by axialities. *Knowledge-Based Systems*. 2010, 23(6), pp. 504–511, doi: 10.1016/j.knosys.2010.03.007.

[22] MIAO D.Q., HU G.R. A heuristic algorithm for reduction of knowledge. *Journal of Computer Research and Development*. 1999, 36(6), pp. 681–684.

[23] POELMANS J., IGNATOV D.I., KUZNETSOV S.O., DEDENE G. Formal concept analysis in knowledge processing: A survey on applications. *Expert Systems with Applications*. 2013, 40(16), pp. 6538–6560, doi: 10.1016/j.eswa.2013.05.009.

[24] QIU G.F., MA J.M., YANG H.Z., ZHANG W.X. A mathematical model for concept granular computing systems. *Science China Information Sciences*. 2010, 53(7), pp. 1397–1408, doi: 10.1007/s11432-010-3092-z.

[25] REN R.S., WEI L. The attribute reductions of three-way concept lattices. *Knowledge-Based Systems*. 2016, 99, pp. 92–102, doi: 10.1016/j.knosys.2016.01.045.

[26] SAMPATH S., SPRENKLE S., GIBSON E., POLLOCK L., GREENWALD A.S. Applying concept analysis to user-session-based testing of web applications. *IEEE Transactions on Software Engineering*. 2007, 33(10), pp. 643–658, doi: 10.1109/TSE.2007.70723.

[27] SHAO M.W., YANG H.Z., WU W.Z. Knowledge reduction in formal fuzzy contexts. *Knowledge-Based Systems*. 2015, 73, pp. 265–275, doi: 10.1016/j.knosys.2014.10.008.

[28] SNELTING G. Reengineering of configurations based on mathematical concept analysis. *ACM Transactions on Software Engineering and Methodology*. 1996, 5(2), pp. 146–189, doi: 10.1145/227607.227613.

[29] SINGH P.K., ASWANI KUMAR Ch. Bipolar fuzzy graph representation of concept lattice. *Information Sciences*. 2014, 288, pp. 437–448, doi: 10.1016/j.ins.2014.07.038.

[30] SINGH P.K., ASWANI KUMAR Ch., LI J. Concepts reduction in formal concept analysis with fuzzy setting using Shannon entropy. *International Journal of Machine Learning and Cybernetics*. 2016, doi: 10.1007/s13042-014-0313-6.

[31] TAN A.H., LI J.J., LIN G.P. Connections between covering-based rough sets and concept lattices. *International Journal of Approximation Reasoning*. 2015, 56, pp. 43–58, doi: 10.1016/j.ijar.2014.09.001.

[32] WANG X.Z. Learning from big data with uncertainty-editorial. *Journal of Intelligent & Fuzzy Systems*. 2015, 28(5), pp. 2329–2330, doi: 10.3233/IFS-141516.

[33] WANG X.Z., Ashfaq R.A.R., Fu A.M. Fuzziness based sample categorization for classifier performance improvement. *Journal of Intelligent & Fuzzy Systems*. 2015, 29(3), pp. 1185–1196, doi: 10.3233/IFS-151729.

[34] WANG X.Z., DONG L.C., YAN J.H. Maximum Ambiguity-Based Sample Selection in Fuzzy Decision Tree Induction. *IEEE Transactions on Knowledge and Data Engineering*. 2012, 24(8), pp. 1491–1505, doi: 10.1109/TKDE.2011.67.

[35] WANG L.D., LIU X.D. Concept analysis via rough set and AFS algebra. *Information Sciences*. 2008, 178(21), pp. 4125–4137, doi: 10.1016/j.ins.2008.07.004.

[36] WANG G.Y., Yu H., YANG D.C. Decision table reduction based on conditional information entropy. *Chinese Journal of Computer*. 2002, 25(7), pp. 759–766.

[37] WANG X., ZHANG W.X. Relations of attribute reduction between object and property oriented concept lattices. *Knowledge-Based Systems*. 2008, 21, pp. 398–403, doi: 10.1016/j.knosys.2008.02.005.

[38] WEI L., QI J.J., ZHANG W.X. Attribute reduction theory of concept lattice based on decision formal contexts. *Science in China Series F: Information Sciences*. 2008, 51(7), pp. 910–923, doi: 10.1007/s11432-008-0067-4.

[39] WILLE R. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts. In: I. RIVAL, ed. *Ordered Sets. Reidel: Dordrecht-Boston*, 1982, pp. 445–470, doi: 10.1007/978-94-009-7798-3_15.

[40] WU W.Z., LEUNG Y., MI J.S. Granular computing and knowledge discovery in formal contexts. *IEEE Transactions on Knowledge and Data Engineering*. 2009, 21(10), pp. 1461–1474, doi: 10.1109/TKDE.2008.223.

[41] YAO Y.Y. Concept lattices in rough set theory. In: S. DICK, L. KURGAN, W. PEDRYCZ, eds. *Proceedings of 23^{rd} International Meeting of the North American Fuzzy Information Processing Society*, Banff, Canada. Washington DC: IEEE Computer society, 2004, pp. 796–801, doi: 10.1109/NAFIPS.2004.1337404.

[42] YAO Y.Y. Information granulation and rough set approximation. *International Journal of Intelligent Systems*. 2001, 16(1), pp. 87–104, doi: 10.1002/1098-111X(200101)16:1<87::AID-INT7>3.0.CO;2-S.

[43] ZADEH L.A. Fuzzy Sets and Information Granularity. In: N. GUPTA, R. RAGADA, R. YAGER, eds. *Advances in Fuzzy Set Theory and Applications*. NorthHolland: Amsterdam, 1979, pp. 3–18.

[44] ZADEH L.A. Towards a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets and Systems*. 1997, 19, pp. 111–127, doi: 10.1016/S0165-0114(97)00077-8.

[45] ZHANG W.X., Ma J.M., FAN S.Q. Variable threshold concept lattices. *Information Sciences*. 2007, 177(22), pp. 4883–4892, doi: 10.1016/j.ins.2007.05.031.

[46] ZHANG W.X., QIU G.F. *Uncertain Decision Making Based on Rough Sets*. Beijing: Tsinghua University Press, 2005.

[47] ZHANG W.X., WEI L., QI J.J. Attribute reduction theory and approach to concept lattice. *Science in China Series F: Information Sciences*. 2005, 48(6), pp. 713–726, doi: 10.1360/122004-104.

[48] ZHI H., LI J. Granule description based on formal concept analysis. *Knowledge-Based Systems*. 2016, 104, pp. 62–73, doi: 10.1016/j.knosys.2016.04.011.