Knowledge-Based Systems 105 (2016) 147–159

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A study on information granularity in formal concept analysis based on concept-bases

Xiangping Kang[a,b,*], Duoqian Miao[a,b]

[a] Department of Computer Science and Technology, Tongji University, Shanghai 201804, China
[b] Key Laboratory of Embedded System and Service Computing (Tongji University), Ministry of Education, Shanghai 201804, China

ARTICLE INFO

Article history:
Received 13 December 2015
Revised 18 March 2016
Accepted 7 May 2016
Available online 9 May 2016

Keywords:
Concept lattice
Granularity
Concept-bases
Concept similarity

ABSTRACT

As one of mature theories, formal concept analysis (FCA) possesses remarkable mathematical properties, but it may generate massive concepts and complicated lattice structure when dealing with large-scale data. With a view to the fact that granular computing (GrC) can significantly lower the difficulty by selecting larger and appropriate granulations when processing large-scale data or solving complicated problems, the paper introduces GrC into FCA, it not only helps to expand the extent and intent of classical concept, but also can effectively reduce the time complexity and space complexity of FCA in knowledge acquisition to some degree. In modeling, concept-base, as a kind of low-level knowledge, plays an important role in the whole process of information granularity. Based on concept-base, attribute granules, object granules and relation granules in formal contexts are studied. Meanwhile, supremum and infimum operations are introduced in the precess of information granularity, whose biggest distinction from traditional models is integrating the structural information of concept lattice. In addition, the paper also probes into reduction, core, and implication rules in granularity formal contexts. Theories and examples verify the reasonability and effectiveness of the conclusions drawn in the paper. In short, the paper not only can be viewed as an effective means for the expansion of FCA, but also is an attempt for the fusion study of the two theories.

1. Introduction

With the increasing popularization of internet technology, "rich data and scarce knowledge" has gradually become a more and more important problem. In the case, how to intelligently and automatically extract potential knowledge from the large-scale data has become one of research hotspots in the current data mining field. Essentially, data mining is a process from the data to information to knowledge, similarly, from the perspective of concept cognition, it can also be understood as the process from the data to lower concepts to higher concepts. Concept cognition is an important characteristic of human brain learning, which is a thinking pattern formed in our minds and mainly focuses on concepts formed through the abstraction and summarization of the common essential characteristics of things. In fact, as one of the most important ways to understand the real world and its regularity, concept cognition is an important foundation of people's complicated thought, and it is also an effective means to express and deduce knowledge.

In philosophy, concept is the thinking unit of human understanding of the objective world and its law, which is composed of two parts, namely intent and extent. In this sense, concepts are essentially the abstract, generalization and induction of the objective world, and the process of generating concepts is essentially a process of optimization and evolution from the perceptual to the rational, and from the phenomenon to the essence, and from the scatter to the system. Meanwhile, in order to adapt to rapid changes in the subjective and objective world, concepts are not only the summary of the understanding of objective things, but also the starting point of the new knowledge, people can deduce new concepts from known ones. If "knowledge system" is compared to a building, then concepts can be understood as core elements of it. Therefore, developing new techniques and methods based on concept thinking will surely contribute to the rapid development of data mining.

If people can use methods in mathematical form to simulate the formation process of concepts and discover the relationship among concepts at different levels, then it will has great significance for data mining and knowledge discovery. Therefore, based

* Corresponding author.
E-mail addresses: tongji_kangxp@sina.com (X. Kang), tongji_miaodq@sina.com (D. Miao).

http://dx.doi.org/10.1016/j.knosys.2016.05.005
0950-7051/

on the philosophical understanding of concepts, German scholar Wille proposed FCA [36] in 1982 from the Brikhoff's lattice theory [1], its formal system can well describe the formation process of concepts by mathematical methods, and can help to stimulate people's mathematical thinking for data analysis and knowledge processing under the concept cognition. In FCA, concepts, concept lattice, Galois connection, et al. play greatly important roles. Namely, any concept can be characterized and described from perspectives of intent and extent, which helps to deepen people's accurate understanding and summary cognizance of actual concepts; Concept lattice, as the core data structure of FCA, can intuitively reflect the generalization and specialization relationships between concepts through Hasse graph; Galois connection is the core theoretical basis of concept lattice. In addition, the generating process of concept lattice is actually a process of objects clustering. In recent years, the research on concept lattice has made a series of important results [15,17,22,23,26,33,43]. For current trends and directions of concept lattice, please refer to reference [28].

In recent years, as a mature theory possessing solid mathematical properties, FCA has drawn more and more attention and been widely applied in various fields such as machine learning, decision analysis, data mining. However, along with the development of research, it is difficult to effectively solve complicated problems just through one single theory. Therefore, many scholars have combined FCA with other theories such as fuzzy set [2,8,12,25,42], rough set [3,9,11,14,21,31,34,35,37], neural network [4], probability theory [7], GrC [8,10,16,29,38], et al., thus greatly expanding the theoretical foundation and application scope of FCA. For example, Burusco et al. brought fuzzy theory into FCA, which can accurately express the uncertain relationship between attributes and objects, thus breaking the binary limitations of classical FCA and helping solve fuzzy and uncertain information in actual applications [2]; Dias et al. combined neural network and FCA and proposed the FCANN [4]; Jiang et al. combined probability theory and FCA, and proposed a new data mining method SPICE [7]; Kang et al. introduced GrC into FCA, and provided a unified model for concept lattice building and rule extraction on the basis of a fuzzy granularity base [8]; The reference [10] introduced FCA and GrC into ontology learning, and presented a unified research model for ontology building, ontology merging and ontology connection based on the domain ontology base in different granulations; Kent discussed the relationship between concept lattice and rough set theory, and presented rough concept analysis that can be viewed as a synthesis of rough set and FCA [11]; Tan et al. studied connections between covering-based rough sets and FCA [31]; Ventos and Soldano presented α Galois lattices based on equivalence classes [32].

FCA, as a data analysis tool, possesses characteristics such as completeness and precision. For a classical concept, even though an object possesses most of attributes of the intent, the object is still not included in the extent of the concept. At the time, it is difficult to manifest the object "possibly" included in concept. Such precision is an advantage of FCA, but it also results in some limitations in processing some specific knowledge. For example, in the process of earthquake prediction, the earth may not always has all characteristics of the earthquake, so experts can only judge that the earthquake possibly happen. However, if all characteristics are identical, then the earthquake may have come about. Since the consequence may be serious, the possibility can not be overlooked. In addition, in some large-scale or complicated data, as a matter of fact, it still deserves attention that such precision of concept lattice often results in a mass of concepts and makes the structure of concept lattice extremely complicated. Aiming at above problems, to better understand and solve problems rather than get lost in unnecessary details, and to better discover potentially valuable knowledge from seemingly irrelevant data, the paper introduces the idea of GrC into FCA, which can, via the unique advantage in the modeling and analysis of large-scale complicated data, lower the "resolution" of knowledge acquisition, and expand the "scale" of knowledge measurement, thus can effectively simplify the complicated concept lattice structure and compact the huge concept scale and prevent some useful information from being buried in massive information.

As a feasible and effective solution for data mining and knowledge reasoning, GrC has become a hotspot research subject in the artificial intelligence field, and a great deal of articles were published. American famous mathematician Zadeh [40] first presented and discussed fuzzy information granularity on the basis of the fuzzy set theory in 1979. In 1997, "granular computing" concept was first formally presented. Some important theoretical findings concerning GrC are shown as follows: Aiming to solve the problem of fuzzy intelligent control through natural language fuzzy reasoning and judgment, Zadeh presented the theory of computing with words [41], and then Thiele proposed the semantic models for investigating computing with words [30], which facilitated the development of the theory of computing with words; Pawlak proposed the rough set theory used for uncertain knowledge modeling, and offered a modeling tool in the case that priori knowledge was incomplete or uncertain [24]; Lin and Yao emphatically described the significance of GrC, that roused people's tremendous interest [18,39]; Zhang et al. presented the GrC model based on quotient space often used in solving complicated problems [44]; Leung and Li described the basic "granule" in an information system with maximal consistent blocks [19]; Hu studied mixed data-oriented neighborhood relation granular computing model and its application [6]; Liang, Qian et al. conducted the systematic research on various uncertain measures, axiomatization of granulation measurement, and rule acquisition in information systems, and studied problems such as incomplete multi-granulation rough set and granulation space structure [20,27].

GrC is a new theory effectively simulating human's thinking and solving complicated problems in the intelligent information processing field. It possesses the unique advantage in the modeling and analysis of large-scale complicated data. No matter from the macro-perspective of cognitive philosophy or from the micro-perspective of information processing, GrC characterized by information granulation, relationships between granules, and granule-based reasoning essentially reflects human's features in solving complicated problems. GrC changed some of our traditional computing concepts in actual applications, making it more scientific, reasonable and operable to deal with problems. For example, when problems are too complicated or solving them requires high cost, the method no longer focuses on some inessential detailed information and takes mathematical exact solutions as the goal, but replaces exact solutions with the feasible satisfactory approximate solutions on the basis of the actual needs so as to achieve the goals of simplifying the problems and enhancing the problem-solving efficiency.

Normally, there are two types of knowledge acquisition methods based on the fusion theory of FCA and GrC. Namely, one type is indirect, which needs data preprocessing, and further using traditional methods to acquire knowledge; While another type is direct [32], which does not need data preprocessing, and can directly deal with original formal contexts. In fact, the former is relatively simple and easy to use, while the latter can completely preserve the original information, which may be more objective than the former.

For instance, for the formal context shown in Table 1(a), indirect methods need data preprocessing, namely, transforming it into a granularity context like Table 1(b), and further using following classical operators (see Definition 1) to acquire knowledge in

Table 1
Some simple formal contexts.

(a)

	a	b	c	d	e
1		×	×	×	
2		×		×	
3	×	×	×		
4	×	×	×	×	×
5	×				×

(b)

	a	b	c	d	e
1		×	×	×	
2		×	•	×	
3	×	×	×	•	•
4	×	×	×	×	×
5	×				×

granularity contexts.

$$A' = \{m \in M \mid gIm, \forall g \in A\}$$

$$B' = \{g \in G \mid gIm, \forall m \in B\}$$

Direct methods directly deal with original formal contexts, such as the method based on operators defined as follows, which are also our research focus at present.

$$A^{+} = \{m \in M \mid \forall g \in A, |m' \cap [g]_R| \geq |[g]_R| \cdot \omega\}$$

$$B^{+} = \{g \in G \mid \forall m \in B, |m' \cap [g]_R| \geq |[g]_R| \cdot \omega\}$$

where R is an equivalence relation on G, and any $[x]_R = \{y \in G | (x, y) \in R\}$ is an equivalence class.

At present, the research on the fusion theory of FCA and GrC is still at an early stage and correlative are rare. In the case, the paper tries to bring GrC into FCA, and proposes an expansion model of FCA based on GrC. Namely, by selecting appropriate and larger granulations not only can help to hide some specific details and lower the difficulty of problems when processing large-scale data or solving complicated problems, but also can help us to discover potentially valuable knowledge from seemingly irrelevant data. In short, the paper not only serves as an expansion of classical FCA, but also offers a new idea for the fusion study of FCA and GrC.

This paper is organized as follows: Section 2 briefly recalls some basic notions of concept lattice; Section 3 defines concept-bases, presents some concept similarity models, and constructs granularity concept-bases by selecting appropriate larger granulations; Section 4 mainly discusses granularity formal contexts; Section 5 presents granularity concept lattices; Section 6 mainly probes into attribute reduction, core and implication rules in granularity formal contexts on the basis of concept-bases; Conclusions and discussions of further work will close the paper in Section 7.

2. Basic notions of concept lattice

This section only offers a brief overview of concept lattice, for more detailed information, please refer to [5].

An order relation on the set L denoted as"$\preceq$", always satisfies reflexivity, antisymmetry, transitivity. In this case, we say $(L, \preceq)$ is an ordered set. And further, we say $s \in L$ is a lower bound of $E \subseteq L$ with $s \preceq q$ for all $q \in E$; An upper bound of E is defined dually. If there exists a single largest element in the set of all lower bounds of E, it is called the infimum of E and is denoted as $\wedge E$; Dually, a single least upper bound is called supremum and is denoted as $\vee E$. For any subset $E \subseteq L$, if there always exist $\wedge E$ and $\vee E$, then we say $(L, \preceq)$ is a complete lattice. In addition, let $x, y \in L$, if $x \preceq y$, we say $[x, y] = \{z \in L | x \preccurlyeq z \preccurlyeq y\}$ is an ordered interval in $(L, \preceq)$.

Normally, the term "formal context" has always been described as a triple, that is, $K = (G, M, I)$, where $I \subseteq G \times M$ is a binary relation between the set G and the set M. In such case, any $g \in G$ is called an object and any $m \in M$ is called an attribute (also called a characteristic), and gIm or $(g, m) \in I$ means the attribute or characteristic m belongs to the object g. Table 2 is a typical formal context, as shown below.

Table 2
A typical formal context.

	a	b	c	d	e	f	g	h
1					×	×	×	×
2	×	×			×	×	×	×
3	×	×	×	×	×			
4	×	×	×	×				
5		×					×	×
6		×	×	×			×	×
7			×	×	×	×		
8				×	×	×		
9			×	×	×	×	×	

Definition 1. In $K = (G, M, I)$, let $A \subseteq G$, $B \subseteq M$, then we define

$$A' = \{m \in M \mid gIm, \forall g \in A\}$$

$$B' = \{g \in G \mid gIm, \forall m \in B\}$$

If $A' = B$ and $B' = A$, then (A, B) is called a concept. The order relationship "$\preceq$" between concepts (A_1, B_1) and (A_2, B_2) is defined as

$$(A_1, B_1) \preccurlyeq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow B_1 \supseteq B_2$$

In fact, the ordered set $(\mathscr{B}(K), \preccurlyeq)$ is a complete lattice, where $\mathscr{B}(K)$ is the set of all concepts. In the case, there are following simple facts

$(A_1, B_1) \wedge (A_2, B_2) = (A_1 \cap A_2, (B_1 \cup B_2)'')$
$(A_1, B_1) \vee (A_2, B_2) = ((A_1 \cup A_2)'', B_1 \cap B_2)$

In addition, for any $g \in G$ and $m \in M$, we say $\gamma g = (g'', g')$ is an object concept and $\mu m = (m', m'')$ is an attribute concept. The set of all object concepts is denoted as $\gamma(G)$, the set of all attribute concepts is denoted as $\mu(M)$.

On the basis of operators in Definition 1, the concept lattice shown in Fig. 2 can be derived from Table 2, the brief process is shown in Fig. 1. In order to make it convenient for formal description, for any object concept (g'', g') and attribute concept (m', m'') in Fig. 2, they are simplified as "g" and "m" separately.

Proposition 1. *In $K = (G, M, I)$, let A, A_1, $A_2 \subseteq G$, B, B_1, $B_2 \subseteq M$, then*

(1) $A_1 \subseteq A_2 \Rightarrow A_2' \subseteq A_1'$ (2) $B_1 \subseteq B_2 \Rightarrow B_2' \subseteq B_1'$
(3) $A \subseteq A''$; $B \subseteq B''$ (4) $A' = A'''$; $B' = B'''$

Proposition 2. *In $K = (G, M, I)$, let $g \in G$, $m \in M$, then*

$(g, m) \in I$, if and only if $\gamma g \preccurlyeq \mu m$

For ease of understanding, in the paper, if the crossing of $g \in G$ row and $m \in M$ column is denoted as "$\diamond$", then we suppose $(g, m) \notin I$; if the crossing of $g \in G$ row and $m \in M$ column is denoted as "•" or "$\times$", then we suppose $(g, m) \in I$.

3. Concept-bases and the granularity of concept-bases

It is known to us that concept lattice is a kind of hierarchical structure model of concepts, and object concepts and attribute concepts, as a kind of basic concepts, are mainly distributed at the upper layer and bottom layer of concept lattice. Essentially, other concepts can be derived from object concepts by means of supremum operation "$\vee$", or derived from attribute concepts by means of infimum operation "$\wedge$", so they play important roles in concept lattice. Therefore, the paper takes the set of object concepts and

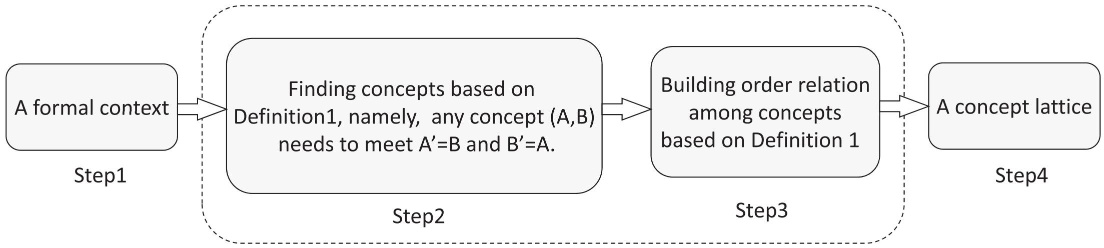

Fig. 1. A brief generation process of concept lattice.

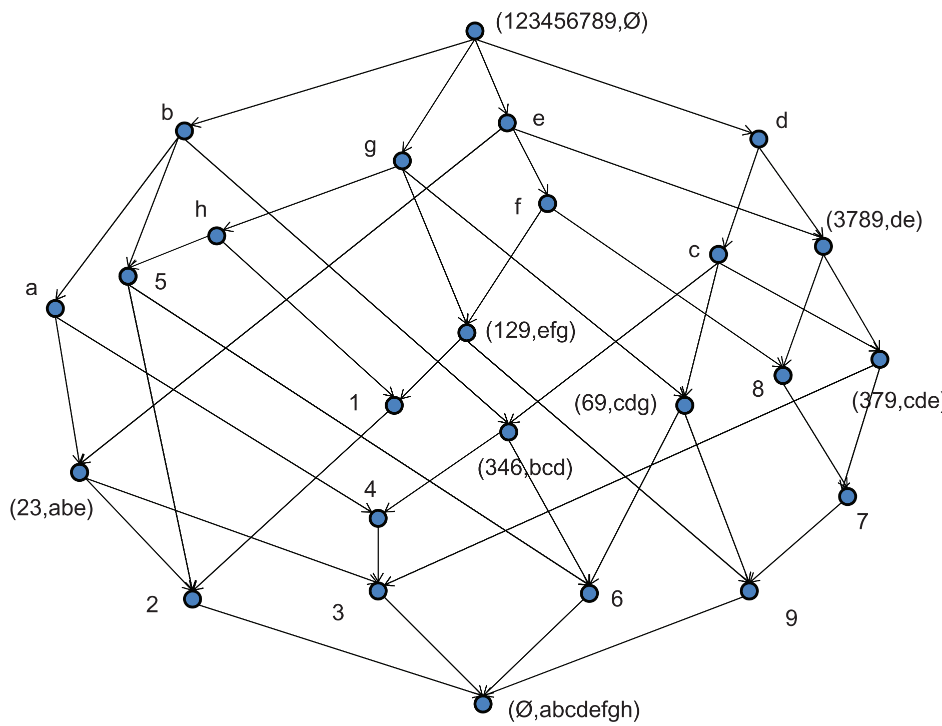

Fig. 2. A concept lattice derived from Table 2.

the set of attributes concept as concept-bases, and proposes a solution to information granularity in FCA based on concept-bases.

Let $(L, \preceq)$ be a concept lattice, $E \subseteq L$, $x \in L$. If there exists $E_1 \subseteq E$ satisfying $x = \vee E_1$, then we say "x" can be derived from E based on the operation "$\vee$". In this case, if all elements in L can be derived from E, then E is called an object concept-base of $(L, \preceq)$; Similarly, if there exists $E_2 \subseteq E$ satisfying $x = \wedge E_2$, then we say "x" can be derived from E based on the operation "$\wedge''$. In this case, if all elements in L can be derived from E, then E is called an attribute concept-base of $(L, \preceq)$.

Proposition 3. *In $K = (G, M, I)$, let (A, B) be a concept, then*

$$\bigwedge_{m \in B} \mu m = (A, B) = \bigvee_{g \in A} \gamma g$$

It is known from the Proposition above that every concept (A, B) can be derived from $\gamma(G)$ based on "$\vee$", and meanwhile can be derived from $\mu(M)$ based on "$\wedge''$. It is obvious that $\gamma(G)$ and $\mu(M)$ are the object concept-base and attribute concept-base of concept lattice $(\mathscr{B}(K), \preccurlyeq)$ respectively.

Theorem 1. *In $(\mathscr{B}(K), \preccurlyeq)$, $\mu(M)$ is an attribute concept-base, and $\gamma(G)$ is an object concept-base.*

We know that the higher time complexity and space complexity of concept lattice generating algorithm have always been the primary obstacle insurmountable in its applications. Especially, when the data scale is larger, most algorithms are still far from being perfect in particular. For instance, when$|M| = n$, lattice nodes amount to 2^n in the worst case. Even in ordinary situations (any object is assumed to have k attributes at most), the upper bound of lattice node number can be as high as $N = 2 + C_n^1 + C_n^2 + \cdots + C_n^k$. Therefore, the paper introduces GrC into FCA, GrC not only helps expand the extent and intent of classical concept, but also effectively reduces the time complexity and space complexity of concept lattice in knowledge acquisition to some degree.

3.1. Concept similarity models

Among traditional concept similarity measurement models, the most common one is characteristic model. One type refers to those meeting symmetry, for instance,

$$sim_1(x, y) = \frac{|B \cap D|}{|B \cup D|}$$

where B and D are intents in concepts x and y respectively. The model above is only on the basis of public characteristics between B and D, and meets symmetry, that is, x and y are similar to each other. It is known to us that intent and extent in any concept are mutually determined, which means that the concept can be determined either by intent or by extent. Therefore, for any concept $x = (A_1, B_1)$ and $y = (A_2, B_2)$, we can describe above similarity model as follows:

$$sim_E(x, y) = |A_1 \cap A_2| \times \frac{1}{|A_1 \cup A_2|}$$

$$sim_I(x, y) = |B_1 \cap B_2| \times \frac{1}{|B_1 \cup B_2|}$$

For instance, for concepts $x = (789, def), y = (379, cde)$ in Table 2, the similarity $sim_E(x, y)$ is

$$sim_E(x, y) = |\{7, 8, 9\} \cap \{3, 7, 9\}| \times \frac{1}{|\{7, 8, 9\} \cup \{3, 7, 9\}|} = 0.5$$

the similarity $sim_I(x, y)$ is

$$sim_I(x, y) = |\{d, e, f\} \cap \{c, d, e\}| \times \frac{1}{|\{d, e, f\} \cup \{c, d, e\}|} = 0.5$$

Essentially, similarity models mentioned-above, no matter on the basis of intents or extents, belong to the characteristic model. Another type is characteristic models stressing asymmetry, for instance:

$$sim_2(x, y) = \frac{|B \cap D|}{|B \cap D| + \alpha \times |B - D| + (1 - \alpha) \times |D - B|}, \quad 0 \leqslant \alpha \leqslant 1$$

where B and D are intents in concepts x and y respectively. $B - D$ denotes characteristic set appearing in B but not in D; $D - B$ denotes characteristic set appearing in D but not in B. The model comprehensively considers the common characteristics and different characteristics between B and D, and assumes that common characteristics affect the similarity more significantly than different characteristics. The parameters α and $1 - \alpha$ can be viewed as weights added to $B - D$ and $D - B$ separately, which help to objectively express the importance of different features and different contributions of $B - D$ and $D - B$ relative to the overall similarity measure. Essentially, α mainly reflects following facts, namely, the contribution of different characteristics is smaller than that of common characteristics in the similarity measure; the influences of different characteristic sets $B - D$ and $D - B$ on similarity maybe not symmetrical, only when $\alpha = 0.5$, the corresponding similarity is symmetrical. For instance, a frequently cited example is that people think the similarity of North Korea relative to China is greater than that of China relative to North Korea.

In the following part, the paper will focus on similarity models similar to $sim_1(x, y)$ because it is simple and unambiguous.

In fact, to better understand and solve problems rather than get lost in unnecessary details of problems in the process of solving problems, we usually hide some specific details so as to obtain their approximate solutions, and the problem can be solved from the overall picture. In virtue of the idea above, the paper constructs a kind of similarity models based on "$\wedge$" and "$\vee$", namely, estimating the similarity between concepts x and y through computing the similarity between $x \wedge y$ and $x \vee y$. Its biggest difference from traditional models is the incorporation of concept lattice's structure information. In fact, such model is constructed on the basis of the following reasonable inferences:

- if $[x, y] \subseteq [v, w]$, then $sim(v, w) \le sim(x, y)$;
- for any concepts $x, y \in [v, w]$, the smaller the interval $[v, w]$ is, the closer $sim(x, y)$ and $sim(v, w)$ are to each other;
- It is assumed that v is the maximum concept smaller than both concepts x and y, and w is the minimum concept bigger than both concepts x and y. It is obvious that among various intervals, $[v, w] = [x \wedge y, x \vee y]$ is the minimum interval including both concepts x and y. Based on above inference, we can approximately estimate the similarity between concepts x and y as the similarity between concepts v and w, namely $sim(x, y) \approx sim(v, w) \approx sim(x \wedge y, x \vee y)$.

Definition 2. Based on above discussion and reasonable inferences, we define similarity models like that: let x and y be concepts, then

$$sim_{LE}(x, y) = sim_E(x \wedge y, x \vee y)$$

$$sim_{LI}(x, y) = sim_I(x \wedge y, x \vee y)$$

For instance, for concepts $x = (789, def), y = (379, cde)$ in Table 2, since $x \wedge y = (79, cdef)$ and $x \vee y = (3789, de)$, the similarity $sim_{LE}(x, y)$ is

$$sim_{LE}(x, y) = |\{7, 9\} \cap \{3, 7, 8, 9\}| \times \frac{1}{|\{7, 9\} \cup \{3, 7, 8, 9\}|} = 0.5$$

the similarity $sim_{LI}(x, y)$ is

$$sim_{LI}(x, y) = |\{d, e\} \cap \{c, d, e, f\}| \times \frac{1}{|\{d, e\} \cup \{c, d, e, f\}|} = 0.5$$

Theorem 2. *In $(\mathscr{B}(K), \preccurlyeq)$, let x and y be concepts, then*

(1) $0 \le sim_{LE}(x, y) \le 1$;
(2) *if* $x = y$, *then* $sim_{LE}(x, y) = 1$;
(3) *if* $x \preceq v \preceq y$ *and* $x \preceq w \preceq y$, *then* $sim_{LE}(x, y) \le sim_{LE}(v, w)$;
(4) *if* $x \preceq z \preceq y$, *then* $sim_{LE}(x, y) \le sim_{LE}(x, z)$.

Proof. The conclusions (1) and (2) can be obtained immediately.

(3) Let $x \wedge y = (A_1, B_1), x \vee y = (A_2, B_2), v \wedge w = (C_1, D_1), v \vee w = (C_2, D_2)$. Then we can see that $x \wedge y \preceq v \wedge w \preceq x \vee y$ and $x \wedge y \preceq v \vee w \preceq x \vee y$ from $x \preceq v \preceq y$ and $x \preceq w \preceq y$, this implies $A_1 \subseteq C_1 \subseteq A_2$ and $A_1 \subseteq C_2 \subseteq A_2$. In this case, we can obtain $C_1 \cup C_2 \subseteq A_1 \cup A_2$ and $A_1 \cap A_2 \subseteq C_1 \cap C_2$. Hence $sim_E(x \wedge y, x \vee y) \le sim_E(v \wedge w, v \vee w)$, that is, $sim_{LE}(x, y) \le sim_{LE}(v, w)$ holds.

(4) Let $x \wedge y = (A_1, B_1), x \vee y = (A_2, B_2), x \wedge z = (C_1, D_1), x \vee z = (C_2, D_2)$. Then we can see that $x \wedge y = x \wedge z \preccurlyeq x \vee y$ and $x \wedge y \preceq x \vee z \preceq x \vee y$ from $x \preceq z \preceq y$, this implies $A_1 = C_1 \subseteq A_2$ and $A_1 \subseteq C_2 \subseteq A_2$. In this case, we can obtain $C_1 \cup C_2 \subseteq A_1 \cup A_2$ and $A_1 \cap A_2 = C_1 \cap C_2$. Hence $sim_E(x \wedge y, x \vee y) \le sim_E(x \wedge z, x \vee z)$, that is, $sim_{LE}(x, y) \le sim_{LE}(x, z)$ holds. □

Theorem 3. *From above discussions, the following statements hold*

(1) *if* $x \wedge y = v \wedge w$ *and* $x \vee y = v \vee w$, *then* $sim_{LE}(x, y) = sim_{LE}(v, w)$;
(2) *for any concepts x and y,* $sim_{LE}(x, y) \le sim_E(x, y)$.

Proof. (1) The conclusion can be obtained immediately.

(2) For any concepts $x = (A_1, B_1)$ and $y = (A_2, B_2)$, let $x \wedge y = (C_1, D_1)$ and $x \vee y = (C_2, D_2)$. Then we can see that $C_1 \subseteq A_1 \subseteq C_2$ and $C_1 \subseteq A_2 \subseteq C_2$ from $x \wedge y \preceq x \preceq x \vee y$ and $x \wedge y \preceq y \preceq x \vee y$, this implies $A_1 \cup A_2 \subseteq C_1 \cup C_2$ and $C_1 \cap C_2 \subseteq A_1 \cap A_2$. Hence $sim_E(x \wedge y, x \vee y) \le sim_E(x, y)$, that is, $sim_{LE}(x, y) \le sim_E(x, y)$ holds. □

Similarly, the similarity model sim_{LI} also meets the properties in above theorems as well. It is self-evident from above theorems that similarity models presented in the paper are feasible.

In fact, sim_{LE} combined with structure information of concept lattice, is a kind of rougher similarity models compared with sim_E. In other words, if the measuring result of sim_E is precise, that of sim_{LE} is approximate. Similarly, compared with sim_I, sim_{LI} is also a kind of rougher similarity models. With a view to the fact that Hasse graph can vividly and succinctly express the structure of concept lattice, we can directly and simply judge which concepts have the same similarities, and which concepts have different similarities on the basis of sim_{LE} or sim_{LI}. For example, for any concepts a_i and a_j in Fig. 3, since the supremum is x, and infimum is y, we can immediately judge that similarity among any two concepts in $\{a_1, a_2, a_3, a_4, a_5\}$ is equal to the similarity between concepts x and

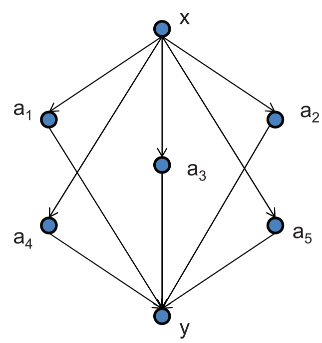

Fig. 3. A concept lattice

Table 3
A fuzzy relation matrix of $\mu(M)$ relative to sim_E.

	μa	μb	μc	μd	μe	μf	μg	μh
μa	1.000							
μb	0.600	1.000						
μc	0.333	0.429	1.000					
μd	0.286	0.375	0.833	1.000				
μe	0.286	0.222	0.375	0.500	1.000			
μf	0.143	0.111	0.250	0.375	0.833	1.000		
μg	0.143	0.429	0.250	0.222	0.375	0.429	1.000	
μh	0.167	0.500	0.125	0.111	0.250	0.286	0.800	1.000

Table 4
A fuzzy relation matrix of $\mu(M)$ relative to sim_{LE}.

	μa	μb	μc	μd	μe	μf	μg	μh
μa	1.000							
μb	0.600	1.000						
μc	0.222	0.333	1.000					
μd	0.222	0.333	0.833	1.000				
μe	0.222	0.222	0.333	0.444	1.000			
μf	0.111	0.111	0.222	0.333	0.833	1.000		
μg	0.111	0.333	0.222	0.222	0.333	0.333	1.000	
μh	0.111	0.333	0.111	0.111	0.222	0.222	0.800	1.000

y. That is, for any concepts a_i and a_j, $sim_{LE}(a_i, a_j) = sim_{LE}(x, y)$ and $sim_{LI}(a_i, a_j) = sim_{LI}(x, y)$, where $i, j = 1, 2, \cdots, 5$.

Based on the discussion above, the paper studies two kinds of similarity models, with the first being similarity models based on concept intent or extent, and the second similarity models based on supremum and infimum. Both of them belong to symmetrical characteristic models in essence.

3.2. The granularity of concept-bases

This section will introduce the transitive closure algorithm in fuzzy clustering analysis, and emphatically make the in-depth analysis of granularity of concept-bases.

In fact, based on one of similarity models presented in the paper, a fuzzy relation matrix can be obtained from the attribute concept-base, such as:

$$F_M = \left(\tilde{r}_{ij}\right)_{|M|\times|M|}, \ \tilde{r}_{ij} = sim(\mu m_i, \mu m_j)$$

where sim represents sim_E or sim_{LE}. For example, on the basis of sim_E, the corresponding fuzzy relation matrix of $\mu(M)$ is shown in Table 3; On the basis of sim_{LE}, the corresponding fuzzy relation matrix of $\mu(M)$ is shown in Table 4. Because of $r_{ij} = r_{ji}$, any r_{ji} is omitted for convenient in above tables.

If F_M satisfies $\tilde{r}_{ii} = 1, \ \tilde{r}_{ij} = \tilde{r}_{ji}$ and $\vee_{k=1,\dots|M|}\tilde{r}_{ik} \wedge \tilde{r}_{kj} \leqslant \tilde{r}_{ij}, \ i, j \in \{1, 2, \dots, |M|\}$, then we say F_M is a fuzzy equivalence relation matrix. We also know that fuzzy relation, especially fuzzy equivalence relation, finds the important application in many fields, such as

Table 5
A fuzzy equivalence relation matrix derived from Table 3.

	μa	μb	μc	μd	μe	μf	μg	μh
μa	1.000							
μb	0.600	1.000						
μc	0.429	0.429	1.000					
μd	0.429	0.429	0.833	1.000				
μe	0.429	0.429	0.500	0.500	1.000			
μf	0.429	0.429	0.500	0.500	0.833	1.000		
μg	0.500	0.500	0.429	0.429	0.429	0.429	1.000	
μh	0.500	0.500	0.429	0.429	0.429	0.429	0.800	1.000

Table 6
A fuzzy equivalence relation matrix derived from Table 4.

	μa	μb	μc	μd	μe	μf	μg	μh
μa	1.000							
μb	0.600	1.000						
μc	0.333	0.333	1.000					
μd	0.333	0.333	0.833	1.000				
μe	0.333	0.333	0.444	0.444	1.000			
μf	0.333	0.333	0.444	0.444	0.833	1.000		
μg	0.333	0.333	0.333	0.333	0.333	0.333	1.000	
μh	0.333	0.333	0.333	0.333	0.333	0.333	0.800	1.000

fuzzy control, approximate reasoning, fuzzy clustering, etc. However, since it is impossible to directly obtain the fuzzy equivalence relation in actual applications, so we usually construct another fuzzy equivalence relation most similar to it in a sense and the most commonly used approach is the fuzzy transitive closure algorithm. Since F_M constructed in the paper is normally reflexive and symmetrical, but not transitive, we can generate fuzzy equivalence relation matrix via the fuzzy transitive closure

$$F_M^+ = F_M^1 \cup F_M^2 \cup \cdots \cup F_M^{n-1}$$

The initial time complexity of transitive closure algorithm is $O(n^3 \log n)$. However, when data is large-scale, the algorithm perhaps can not finish generating fuzzy equivalence relation matrix in the limited time. Lee et al.[13] proposed an algorithm with $O(n^2)$. In view of its low time complexity, we use it to compute the transitive closure F_M^+. For example, by using the transitive closure algorithm, a fuzzy equivalence relation matrix shown in Table 5 can be derived from Table 3; In the same way, we can obtain Table 6 from Table 4.

And further, we can obtain the following $\sigma-$cut equivalence relation matrix $F_M^+(\sigma)$ by introduce the parameter $\sigma \in [0, 1]$.

$$F_M^+(\sigma) = \left(r_{ij}\right)_{|M|\times|M|}, \ \text{where } r_{ij} = \begin{cases} 1 & \tilde{r}_{ij} \geqslant \sigma \\ 0 & \tilde{r}_{ij} < \sigma \end{cases}$$

In fact, $F_M^+(\sigma)$ is an equivalence relation, and $\mu(M)/F_M^+(\sigma) = \{P_1, \cdots P_l\}$ is a partition of $\mu(M)$, where P_i is an equivalence class. Obviously, the attribute concept-base can be granulated into several equivalence classes. In this case, we define

$$\rho_M(\sigma) = \frac{1}{|M|^2} \times \sum_{i=1}^{l} |P_i|^2$$

then we say $\rho_M(\sigma)$ is the granulation of $F_M^+(\sigma)$. Obviously, the bigger σ is, the smaller $\rho_M(\sigma)$ is, and vice versa. In fact, a bigger granulation can help hide some partial details, thus making it convenient for us to view and understand the whole problem comprehensively. Similarly, from the object concept-base $\gamma(G)$, we can obtain $F_G^+(\sigma)$ on the basis of sim_I or sim_{LI}, which will not be detailed here again.

4. Granularity formal contexts

On the basis of discussions in above section, an attribute granule $[m]$ with granulation $\rho_M(\sigma)$ is defined as:

$$[m] = \{n \in M | (\mu m, \mu n) \in F_M^+(\sigma)\}$$

the set of all attribute granules is denoted as M_σ; Correspondingly, an object granule $[g]$ with granulation $\rho_G(\sigma)$ is defined as:

$$[g] = \{h \in G | (\gamma g, \gamma h) \in F_G^+(\sigma)\}$$

the set of all object granules is denoted as G_σ. And further, for any $[m] \in M_\sigma$ and $[g] \in G_\sigma$, we say $[g] \times [m]$ is a relation granule with granulation $\rho_M(\sigma)$.

For example, when $\sigma = 0.6$, then the granularity result of M relative to sim_{LE} or sim_E is

$$M_\sigma = \{[a], [c], [e], [g]\}$$

where $[a] = \{a, b\}$, $[c] = \{c, d\}$, $[e] = \{e, f\}$, $[g] = \{g, h\}$, the corresponding granulation is $\rho_M(\sigma) = 0.25$; Similarly, the granularity result of G relative to sim_{LI} or sim_I with $\sigma = 0.6$ is

$$G_\sigma = \{[1], [3], [5], [7]\}$$

where $[1] = \{1, 2\}$, $[3] = \{3, 4\}$, $[5] = \{5, 6\}$, $[7] = \{7, 8, 9\}$, the corresponding granulation is $\rho_G(\sigma) = 0.31$.

Definition 3. In $K = (G, M, I)$, we say $K_\sigma = (G_\sigma, M_\sigma, I_\sigma)$ is a granularity formal context, where $I_\sigma \subseteq G_\sigma \times M_\sigma$. In the case, for any relation granule $[g] \times [m]$, either $([g], [m]) \in I_\sigma$ holds or $([g], [m]) \notin I_\sigma$ holds.

In above definition, the paper offers two different types of judging rules concerning whether the relation granule $[g] \times [m]$ satisfies $([g], [m]) \in I_\sigma$, namely judging methods based on confidence function ϕ_σ and based on supremum and infimum separately. Their biggest difference is that the former is the judging method based on binary relation I while the latter is incorporating the structure information of concept lattice.

Definition 4. For any relation granule $[g] \times [m]$, the reliability of $([g], [m]) \in I_\sigma$ is defined as

$$\phi_\sigma([g], [m]) = \frac{1}{|[g]| \times |[m]|} \times |\pi([g], [m])|$$

where $\pi([g], [m])$ is defined as

$$\pi([g], [m]) = \{(h, n) | \gamma h \preccurlyeq \mu n, h \in [g], n \in [m]\}$$

Obviously, the bigger $\phi_\sigma([g], [m])$ is, the higher the reliability of $([g], [m]) \in I_\sigma$ is, and vice versa. Therefore, we introduce parameter $\varpi \in [0, 1]$, view it as the threshold value for judging whether $([g], [m]) \in I_\sigma$ is reliable, and present the following criterion.

Criterion 1. In (G, M, I), for any relation granule $[g] \times [m]$, if $\phi_\sigma([g], [m]) \geq \varpi$, then $([g], [m]) \in I_\sigma$; if $\phi_\sigma([g], [m]) < \varpi$, then $([g], [m]) \notin I_\sigma$.

In Criterion 1, based on following principles, users can adjust the parameter ϖ so as to meet the actual needs. How to set ϖ reasonably, the paper gives following suggestions

(1) when ϖ is bigger, it will be unfavorable for us to discover potentially information from seemingly irrelevant data. For instance, for the relation granule $[g] \times [m]$ shown in Table 7(a), it is obvious that $([g], [m]) \in I_\sigma$ is more reasonable than $([g], [m]) \notin I_\sigma$. However, if people sets $\varpi = 1$, then the conclusion $([g], [m]) \in I_\sigma$ will not be obtained.

(2) when $\varpi < 0.5$, then corresponding results may not be able to provide a scientific basis for further data modeling. For instance, for the relation granule $[g] \times [m]$ shown in Table 7(b), it is obvious that $([g], [m]) \notin I_\sigma$ is more reasonable than $([g], [m]) \in I_\sigma$. However, if people sets $\varpi < 0.5$, then the conclusion $([g], [m]) \notin I_\sigma$ may not be obtained.

Table 7
Some relation granules.

(a)

	×	×	×	×	×
×	×	×	×	×	×
×	×	×	×	×	×
×	×	×	×	×	×
×	×	×	×	×	×
×	×	×	×	×	

(b)

	×		×		
					×
×			×		
		×			
×				×	
		×			×

Table 8
A granularity context derived from Table 2.

	a	b	c	d	e	f	g	h
1	•	•			×	×	×	×
2	×	×			×	×	×	×
3	×	×	×	×	◇			
4	×	×	×	×				
5	•	×	•	•			×	×
6	•	×	×	×			×	×
7			×	×	×	×		
8			•	×	×	×		
9			×	×	×	×	◇	

A granularity formal context can be derived from Table 2 on the basis of Criterion 1, which is shown in Table 8 with $\sigma = 0.6$ and $\varpi = 0.5$.

Criterion 2. In (G, M, I), if $\wedge\gamma([g]) \npreceq \vee\mu([m])$, then $([g], [m]) \notin I_\sigma$; if $\wedge\gamma([g]) \preceq \vee\mu([m])$, then $([g], [m]) \in I_\sigma$.

Theorem 4. *In (G, M, I), let $[g] \times [m]$ be relation granule, then*

(1) *if $\wedge\gamma([g]) \npreceq \vee\mu([m])$, then $\phi_\sigma([g], [m]) = 0$;*
(2) *if $\phi_\sigma([g], [m]) > 0$, then $\wedge\gamma([g]) \preceq \vee\mu([m])$.*

Proof. (1) If $\wedge\gamma([g]) \npreceq \vee\mu([m])$, then we suppose $\phi_\sigma([g], [m]) \neq 0$, namely, there exists $(g_1, m_1) \in [g] \times [m]$ with $(g_1, m_1) \in I$. This implies $\gamma g_1 \preceq \mu m_1$ by Proposition 2. And further together with $\wedge\gamma([g]) \preceq \gamma g_1$ and $\mu m_1 \preceq \vee\mu([m])$, $\wedge\gamma([g]) \preceq \vee\mu([m])$ can be obtained, which contradicts with $\wedge\gamma([g]) \npreceq \vee\mu([m])$. Hence the conclusion (1) holds.

(2) If $\phi_\sigma([g], [m]) > 0$, there must exist $(g_1, m_1) \in [g] \times [m]$ with $(g_1, m_1) \in I$. In the case, $\gamma g_1 \preceq \mu m_1$ can be obtained by Proposition 2. And further we can see $\wedge\gamma([g]) \preceq \vee\mu([m])$ from $\wedge\gamma([g]) \preceq \gamma g_1$ and $\mu m_1 \preceq \vee\mu([m])$. Hence the conclusion (2) holds. □

In Theorem 4, we can see Criterion 2 is essentially a special case of Criterion 1, and the judging condition in Criterion 2 is clearly weaker than that in Criterion 1, so the corresponding judging result is not reasonable enough. Therefore, Criterion 1 is relatively better. In the case, the paper points out Criterion 3 further.

Criterion 3. In (G, M, I), when $|[g]| = 1$ or $|[m]| = 1$, whether the granule $[g] \times [m]$ meets $([g], [m]) \in I_\sigma$ can be judged by Criterion 2; when $|[g]| \geq 2$ and $|[m]| \geq 2$, then

- for any $g_1, g_2 \in [g]$ and $m_1, m_2 \in [m]$, if

  $$g_1 \neq g_2,\ m_1 \neq m_2 \text{ and } \gamma g_1 \wedge \gamma g_2 \preccurlyeq \mu m_1 \vee \mu m_2$$

  then $([g], [m]) \in I_\sigma$;
- if there exist $g_1, g_2 \in [g]$ and $m_1, m_2 \in [m]$ satisfying

  $$g_1 \neq g_2,\ m_1 \neq m_2 \text{ and } \gamma g_1 \wedge \gamma g_2 \npreceq \mu m_1 \vee \mu m_2$$

  then $([g], [m]) \notin I_\sigma$.

For any $g_1, g_2 \in [g]$ and $m_1, m_2 \in [m]$, because of $\wedge\gamma([g]) \preceq (\gamma g_1 \wedge \gamma g_2) \preceq (\mu m_1 \vee \mu m_2) \preceq \vee\mu([m])$, it is obvious that the

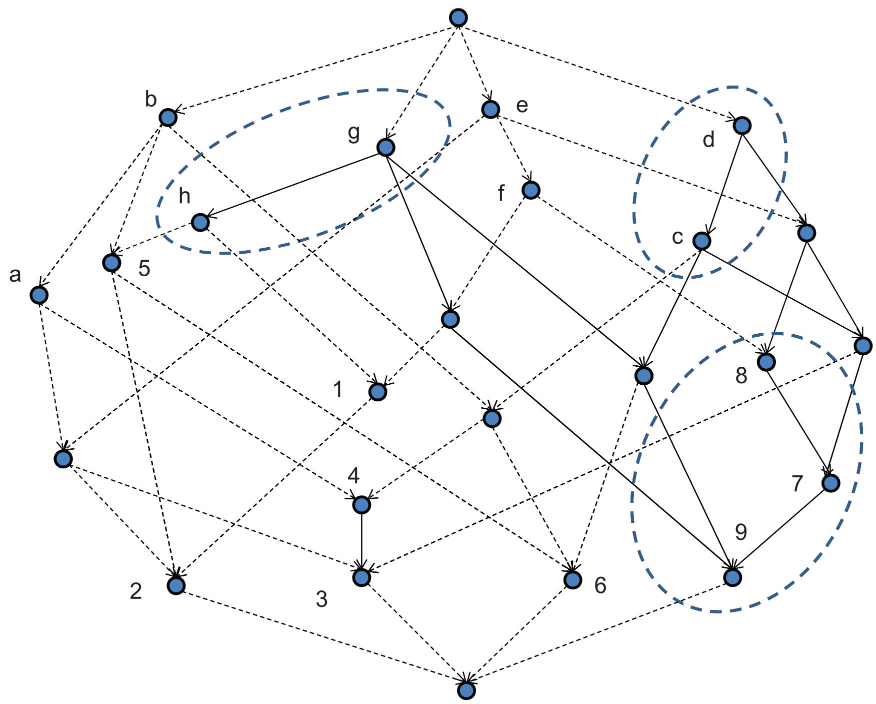

Fig. 4. A concept lattice derived from Table 2.

Table 9
A formal context derived from Table 2.

	a	*b*	*c*	*d*	*e*	*f*	*g*	*h*
1	•	•			×	×	×	×
2	×	×			×	×	×	×
3	×	×	×	×	×	•		
4	×	×	×	×	•	•		
5	•	×	•	•			×	×
6	•	×	×	×			×	×
7			×	×	×	×		
8			•	×	×	×		
9			×	×	×	×	◇	

judging condition $\gamma g_1 \wedge \gamma g_2 \preceq \mu m_1 \vee \mu m_2$ in Criterion 3 is more stronger than that of $\wedge\gamma([g]) \preceq \vee\mu([m])$ in Criterion 2. Therefore, the judging result deduced from Criterion 3 is more reasonable.

A granularity formal context can be derived from Table 2 on the basis of Criterion 3, which is equivalent to the formal context shown in Table 9 with $\sigma = 0.6$.

For example, since $\gamma 7 \wedge \gamma 8 \preceq \mu c \vee \mu d$, $\gamma 7 \wedge \gamma 9 \preceq \mu c \vee \mu d$, $\gamma 8 \wedge \gamma 9 \preceq \mu c \vee \mu d$ by referring Fig. 4, we can see that $([7], [c]) \in I_\sigma$. In addition, $([7], [g]) \notin I_\sigma$ can be derived from $\gamma 7 \wedge \gamma 8 \npreceq \mu g \vee \mu h$, $\gamma 7 \wedge \gamma 9 \preceq \mu g \vee \mu h$, $\gamma 8 \wedge \gamma 9 \preceq \mu g \vee \mu h$ by referring Fig. 4.

Criterion 2 and Criterion 3 based on operations $\vee$ and $\wedge$ can particularly resort to Hasse graph to directly and simply judge whether $[g] \times [m]$ meets $([g], [m]) \in I_\sigma$, whose biggest distinction from traditional models is integrating the structural information of concept lattice. Maybe Criterion 2 or Criterion 3 is not more reasonable than other methods, but it provides a new way of thinking for the related research.

Kang et al. have introduced equivalence relations into M, and studied the similar problem [8]. In this literature, for any attribute granule $[m]$ and object g, a judging criterion on whether $(g, [m]) \in J_\delta$ is essentially defined as

if $\exists m_1 \in [m]$ satisfying $(g, m_1) \in I$, then $(g, [m]) \in J_\delta$.

where (G, M_δ, J_δ) is the granularity context to be solved for. The criterion is equivalent to the following statement

- In (G, M, I), for any relation granule $g \times [m]$, if $\phi_\sigma(g, [m]) > 0$, then $(g, [m]) \in I_\sigma$; if $\phi_\sigma(g, [m]) = 0$, then $(g, [m]) \notin I_\sigma$.

Essentially, above criterion is only a special case of Criterion 1. Since its judging condition is clearly weaker than that in Criterion 1, it is obvious that Criterion 1 is more reasonable.

In addition, the literature [5] provides another way for getting granularity context, that is, as a granularity context of (G, M, I), (G, M, J) needs to meet following conditions

- $I \subseteq J$;
- for every object $g \in G$, g' in (G, M, J) is an intent of (G, M, I);
- for every attribute $m \in M$, m' in (G, M, J) is an extent of (G, M, I).

For instance, Table 1(b) is just a granularity context of Table 1(a). Advantages of above method are listed as follows: there is close relationship between granularity contexts and original contexts, namely, intents in (G, M, J) must be intents in (G, M, I), and extents in (G, M, J) must be extents in (G, M, I); there is close relationship between granularity lattices and original lattices. For instance, the relationship between the concept lattice derived from Table 1(a) and its granularity concept lattice derived from Table 2(b) are shown in Fig. 5. Although there are many advantages, constraint conditions in above method are too many to meet reality applications. Comparatively, methods presented in the paper may be not satisfy above conditions, but they are relatively easier and more practical.

5. Granularity concept lattices

With the strong algebraic structure, concept lattice can accurately display the inheritance relationship among concept nodes, so it is suitable to be used as fundamental data structure of rule discovery for discovering rule-based knowledge. However, when dealing with the large-scale data, FCA may generate massive concepts and complicated lattice structure, and accordingly encounter the tremendous limitations in actual applications. With a view to

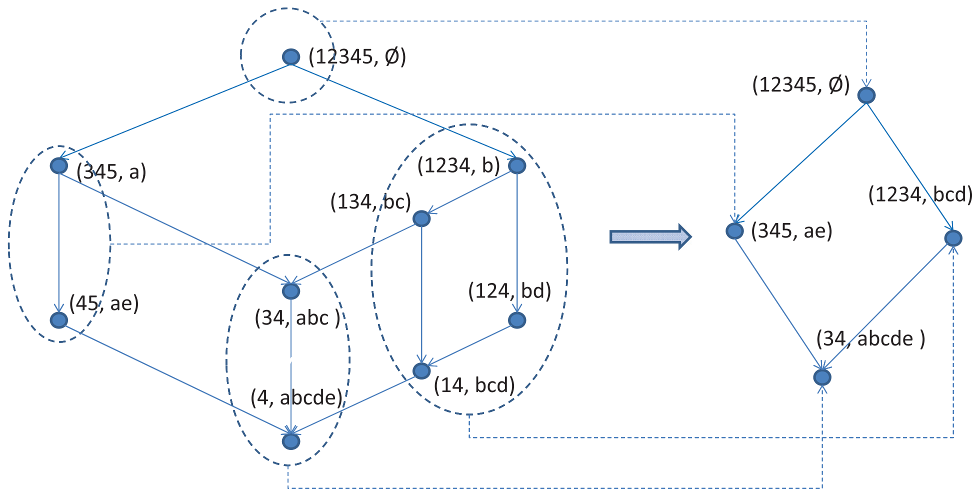

Fig. 5. A concept lattice and its granularity concept lattice.

the fact that GrC itself can simulate human's intelligence, can significantly lower the difficulty of problems by selecting appropriate granulations, thus providing a feasible, effective solution to knowledge mining and knowledge reasoning. Therefore, the section mainly probes into granularity concept lattices in formal contexts.

In $K_\sigma = (G_\sigma, M_\sigma, I_\sigma)$, let $A_\sigma \subseteq G_\sigma$, then

$$A'_\sigma = \{[m] \in M_\sigma \mid ([g], [m]) \in I_\sigma, \forall [g] \in A_\sigma\}$$

Correspondingly, let $B_\sigma \subseteq M_\sigma$, then

$$B'_\sigma = \{[g] \in G_\sigma \mid ([g], [m]) \in I_\sigma, \forall [m] \in B_\sigma\}$$

If $A'_\sigma = B_\sigma$ and $B'_\sigma = A_\sigma$, then (A_σ, B_σ) is called a granularity concept. The order relation "$\preceq_\sigma$" between granularity concepts (A_σ, B_σ) and (C_σ, D_σ) is defined as

$$(A_\sigma, B_\sigma) \preccurlyeq_\sigma (C_\sigma, D_\sigma) \Leftrightarrow A_\sigma \subseteq C_\sigma \Leftrightarrow D_\sigma \supseteq B_\sigma$$

Obviously, $(\mathscr{B}(K_\sigma), \preccurlyeq_\sigma)$ is a complete lattice, where $\mathscr{B}(K_\sigma)$ is the set of all granularity concepts.

In $(G_\sigma, M_\sigma, I_\sigma)$, for any $[g] \in G_\sigma$, $[m] \in M_\sigma$, we say $\widetilde{\gamma}[g] = ([g]'', [g]')$ is an object granularity concept, and $\widetilde{\mu}[m] = ([m]', [m]'')$ is an attribute granularity concept. It is obvious $\widetilde{\gamma}(G_\sigma)$ and $\widetilde{\mu}(M_\sigma)$ are concept-bases of $(\mathscr{B}(K_\sigma), \preccurlyeq_\sigma)$, where the set of all object granularity concepts is denoted as $\widetilde{\gamma}(G_\sigma)$, the set of all attribute granularity concepts is denoted as $\widetilde{\mu}(M_\sigma)$.

Definition 5. Based on discussions above, the lattice $(\mathscr{B}(K_\sigma), \preccurlyeq_\sigma)$ is called a granularity concept lattice of $(\mathscr{B}(K), \preccurlyeq)$.

For example, granularity concept lattices with respect to Table 8 and Table 9, are shown in Fig. 6 and Fig. 7 separately. In essence, the method based on Criterion 1 is an expansion of the one in [8]. For further revealing the similarity and differences between them, the paper takes a practical example as application background, which is shown as follows.

The context shown in Table 10 is about websites and their subjects. The set of objects G is composed of website 1, website 2, ..., website 12 denoting some websites, the set of attributes M is composed of Financing, Economic, ..., Education denoting subjects in websites. If website i includes subject j, this is denoted as $(i, j) \in I$, which is shown in table by " × ". Fig. 8 is the classical concept lattice with respect to Table 10, since it is too complicated and too large to show, we only show it partially. Table 11 and Table 12 are granularity contexts of Table 10, which are results based on [8] and Criterion 1 defined in the paper separately. Corresponding granularity concept lattices are shown in Figure 9 and Fig. 10.

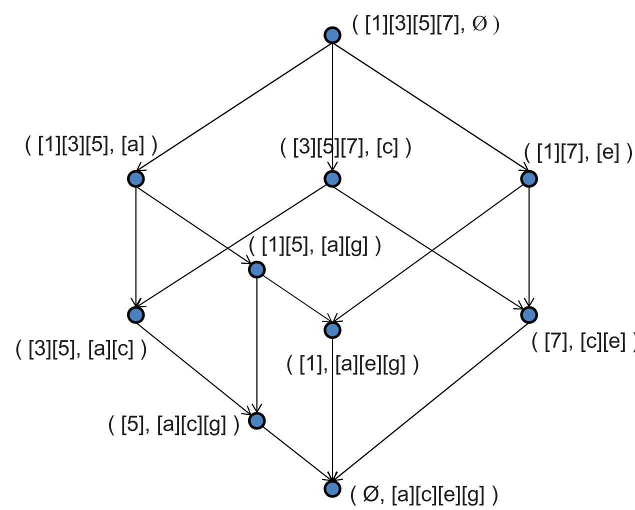

Fig. 6. A granularity concept lattice with respect to Table 8.

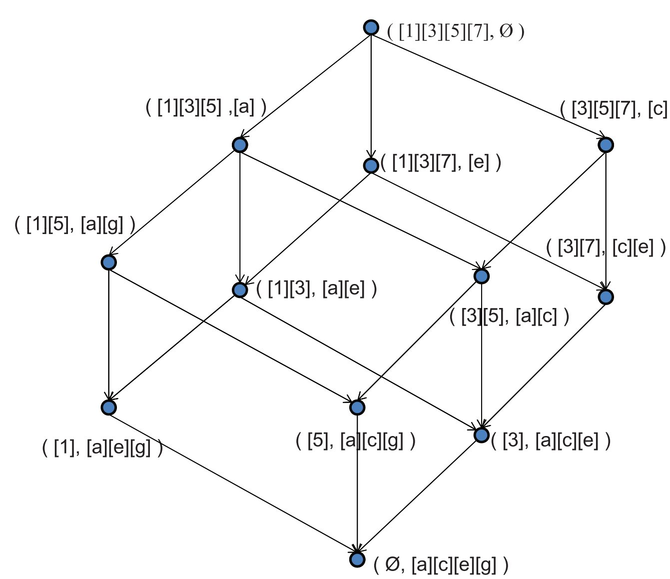

Fig. 7. A granularity concept lattice with respect to Table 9.

Table 10
A formal context about websites and their subjects

	1	2	3	4	5	6	7	8	9	10	11	12
Financing	×	×	×	×	×	×	×					
Economic	×		×	×	×	×						
Stock certification	×	×	×		×	×						
Foundation	×	×	×	×	×							
History	×	×	×			×	×	×	×	×	×	×
Record	×	×	×	×	×	×	×	×	×		×	
Literature	×									×	×	×
Culture		×								×	×	×
Education			×							×	×	×

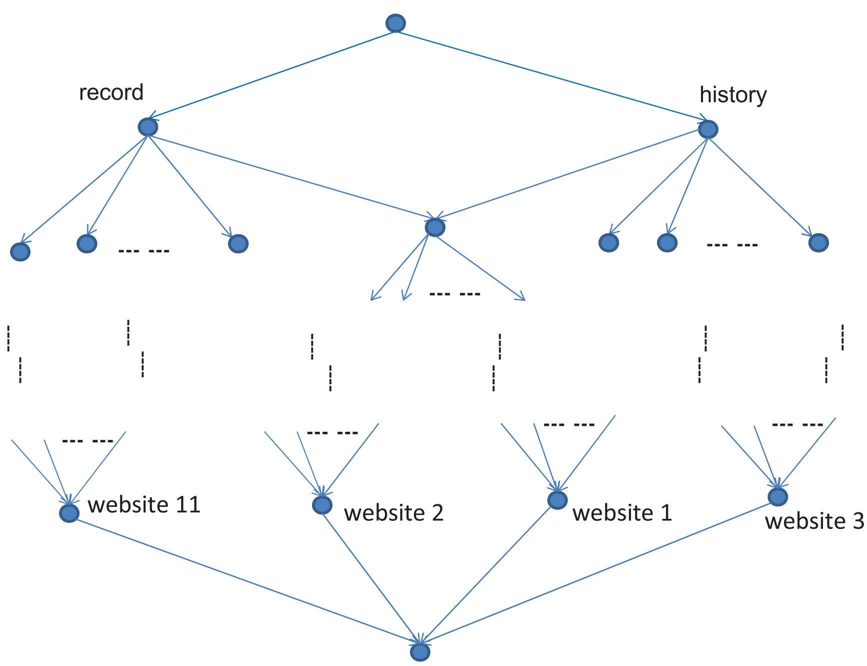

Fig. 8. A classical concept lattice with respect to Table 10.

Table 11
A granularity context derived from Table 10.

	1	2	3	4	5	6	7	8	9	10	11	12
Financing	×	×	×	×	×	×	×					
Economic	×	•	×	×	×	×	•					
Stock certification	×	×	×	•	×	×	•					
Foundation	×	×	×	×	×	•	•					
History	×	×	×	•	•	×	×	×	×	×	×	×
Record	×	×	×	×	×	×	×	×	×	•	×	•
Literature	×	•	•							×	×	×
Culture	•	×	•							×	×	×
Education	•	•	×							×	×	×

Table 12
A granularity context derived from Table 10.

	1	2	3	4	5	6	7	8	9	10	11	12
Financing	×	×	×	×	×	×	◇					
Economic	×	•	×	×	×	×						
Stock certification	×	×	×	•	×	×						
Foundation	×	×	×	×	×	•						
History	×	×	×	•	•	×	×	×	×	×	×	×
Record	×	×	×	×	×	×	×	×	×	•	×	•
Literature	◇									×	×	×
Culture		◇								×	×	×
Education			◇							×	×	×

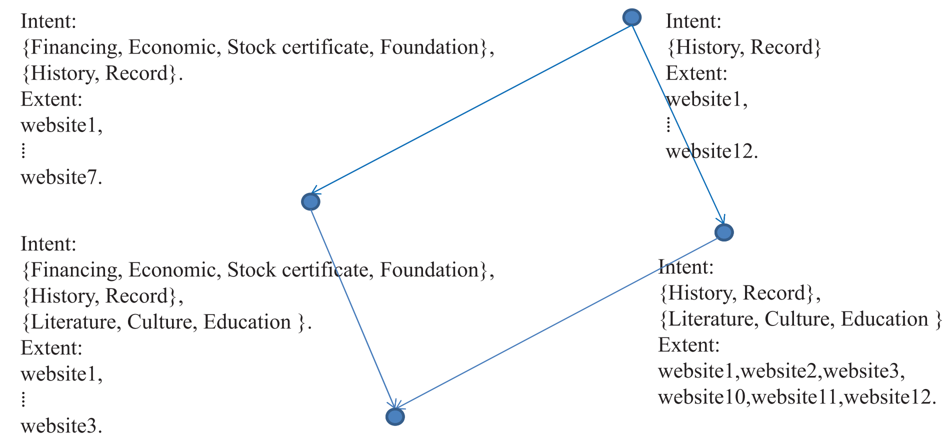

Fig. 9. A granularity concept lattice with respect to Table 11.

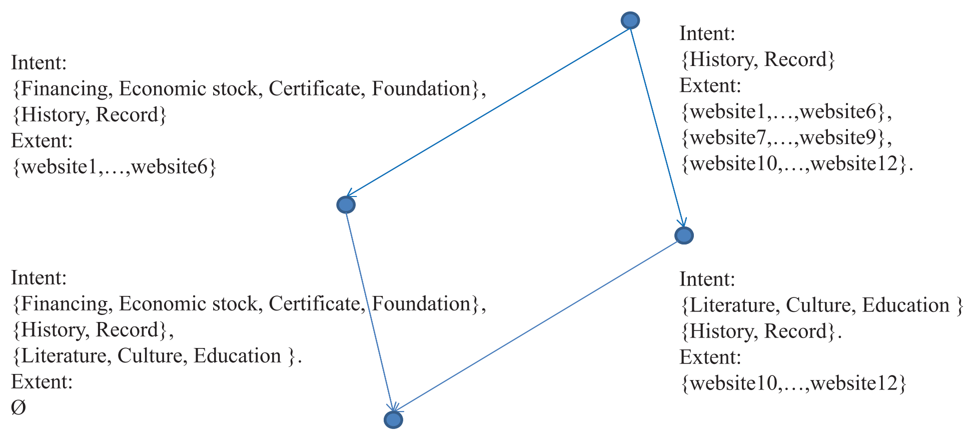

Fig. 10. A granularity concept lattice with respect to Table 12.

In Table 10, when $\sigma = 0.6$ and $\varpi = 0.6$, then all attributes can be classified into {Financing, Economic, Stock certificate, Foundation}, {History, Record} and {Literature, Culture, Education} by Criterion 1; Meanwhile, all objects can be classified into {website1,…,website6}, {website7,…,website9} and {website10, …, website12}. In addition, based on the method in [8], all attributes can also be classified into same granules.

From the actual example above, we can see that there are some similarities and some slight differences relative to [8]. **Similarities:**they all help to compress the scale of structure of concept lattice and reduce the number of concepts by hiding some specific details. For instance, Fig. 9 or Fig. 10 has relatively simpler structure and smaller nodes than Fig. 8; they all help to expand the extent and intent of classical concept. For instance, as approximate concepts of classical concepts, any granularity concept in Fig. 9 and Fig. 10 contain more objects and attributes than the one in Fig. 8; they all introduce equivalence relations and parameters into FCA. **Differences:** by comparing with [8], the paper introduces equivalence relations into both G and M rather than only M, so it is more conducive to thorough expanding classical FCA from the respective of GrC; the judging condition in [8] is clearly weaker than that in Criterion 1. For instance, for any subject and website, if the website includes the subject in Table 12, then there is the same result in Table 11. Since Criterion 1 possesses stronger condition, it is more reasonable relatively to some certain extent.

In addition, Criterion 2 and Criterion 3 based on operations $\vee$ and $\wedge$ can particularly resort to Hasse graph to directly and simply judge whether $[g] \times [m]$ meets $([g], [m]) \in I_\sigma$, whose biggest distinction from traditional models is integrating the structural information of concept lattice. Maybe they are not more reasonable than other methods, but it provides a new way of thinking for the related research.

6. The knowledge acquisition in granularity formal contexts

This section, by means of concept-bases $\widetilde{\gamma}(G_\sigma)$ and $\widetilde{\mu}(M_\sigma)$, mainly probes into attribute reduction, core and implication rules in $K_\sigma = (G_\sigma, M_\sigma, I_\sigma)$.

In $K = (G, M, I)$, let $C \subseteq B \subseteq M$, if C is the minimal subset satisfies $B' = C'$, then we say C is a reduction of B; Let $m \in B \subseteq M$, if $B' \neq (B - m)'$, then we say m is indispensable, the set of all indispensable attributes in B is called the core of B, which is denoted as $core(B)$; let $B, C \subseteq M$, if $B' \subseteq C'$, then we say $B \rightarrow C$ is an implication rule.

Theorem 5. *Let $[m] \in B_\sigma \subseteq M_\sigma$, then $[m] \in core(B_\sigma)$, if*

$$\{[s] \mid \widetilde{\gamma}[s] \preccurlyeq_\sigma \widetilde{\mu}[a], \forall [a] \in B_\sigma\} \neq \{[g] \mid \widetilde{\gamma}[g] \preccurlyeq_\sigma \widetilde{\mu}[b], \forall [b] \in (B_\sigma - [m])\}$$

Proof. From Proposition 2 we can see that $([g], [m]) \in I_\sigma \Leftrightarrow \widetilde{\gamma}[g] \preccurlyeq_\sigma \widetilde{\mu}[m]$. And further, the condition mentioned in the theorem is equivalent to $B'_\sigma \neq (B_\sigma - [m])'$. Hence, $[m] \in core(B_\sigma)$ holds. □

Theorem 6. *Let $C_\sigma \subseteq B_\sigma \subseteq M_\sigma$, then C_σ is a reduction of B_σ, if C_σ is the minimal subset of B_σ satisfying the condition*

$$\{[h] \mid \widetilde{\gamma}[h] \preccurlyeq_\sigma \widetilde{\mu}[m], \forall [m] \in C_\sigma\} = \{[g] \mid \widetilde{\gamma}[g] \preccurlyeq_\sigma \widetilde{\mu}[n], \forall [n] \in B_\sigma\}$$

Proof. For any $[g] \in G_\sigma$, $[m] \in M_\sigma$, $([g], [m]) \in I_\sigma \Leftrightarrow \widetilde{\gamma}[g] \preccurlyeq_\sigma \widetilde{\mu}[m]$ can be deduced from Proposition 2. And further, we can obtain $C'_\sigma = B'_\sigma$. Hence, C_σ is a reduction of B_σ. □

Theorem 7. *Let* C_σ, $B_\sigma \subseteq M_\sigma$, *then* $B_\sigma \to C_\sigma$, *if*

$$\{[h] \mid \widetilde{\gamma}[h] \preccurlyeq_\sigma \widetilde{\mu}[m], \forall [m] \in B_\sigma\} \subseteq \{[g] \mid \widetilde{\gamma}[g] \preccurlyeq_\sigma \widetilde{\mu}[n], \forall [n] \in C_\sigma\}$$

Proof. By means of $([g], [m]) \in I_\sigma \Leftrightarrow \widetilde{\gamma}[g] \preccurlyeq_\sigma \widetilde{\mu}[m]$ deduced from Proposition 2, $B'_\sigma \subseteq C'_\sigma$ can be obtained. Hence, $B_\sigma \to C_\sigma$ holds. □

7. Conclusions

The paper tries to bring GrC into FCA, and proposes an expansion model of FCA based on GrC, which helps to hide some specific details and lower the difficulty of problems, and also helps to discover valuable knowledge from seemingly irrelevant data through the expansion of intent and extent of the classical concept. In modeling, the paper defines concept-bases, discusses the granularity of concept-bases, and further studies granularity formal contexts and granularity concept lattices. In fact, concept-bases, as a kind of low-level knowledge, play important roles in the whole data modeling.

To better understand and solve problems rather than get lost in unnecessary details of problems in the process of solving problems on the basis of GrC, we usually abstract and simplify problems so as to obtain their approximate solutions. In virtue of the idea above, the paper constructs a kind of concept similarity models based on supremum and infimum, namely estimating the similarity between concepts x and y through computing the similarity between $x \wedge y$ and $x \vee y$. Its biggest difference from traditional models is the incorporation of concept lattice's structure information. Since Hasse graph can vividly and succinctly manifest the structure of concept lattice, we can directly and simply judge which concepts have the same similarities, and which concepts have different similarities in the model through Hasse graph. In addition, the paper also presents other concept similarity models, which are essentially characteristic models.

Concerning whether the relation granule $[g] \times [m]$ meets $([g], [m]) \in I_\sigma$, the paper offers the judging methods based on reliability function ϕ_σ and based on operators $\vee$ and $\wedge$. Their biggest difference is that the former is the judging method based on the binary relation I while the latter incorporates the concept lattice's structure information. As for the judging method based on operators $\vee$ and $\wedge$, in particular, we can resort to Hasse graph to directly and simply judge whether $[g] \times [m]$ meets $([g], [m]) \in I_\sigma$.

In the end, the paper proposes granularity concept lattice, and also emphatically probes into attribute reduction, core, and implication rules in granularity formal contexts, and offers solutions based on concept-bases. In short, the introduction of GrC into FCA study can yet be regarded as an effective means for the expansion of FCA. Theories and examples verify the reasonability and effectiveness of conclusions drawn in the paper. Although FCA is an effective tool for data analysis, it still has many limitations in dealing with large-scale data and complicated knowledge discovery tasks. The research in the paper is just a tentative move, and further research remains to be carried out on the fusion theory of GrC and FCA.

Acknowledgements

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the manuscript. This work was supported by the National Postdoctoral Science Foundation of China (No. 2014M560352), the National Natural Science Foundation of China (Nos. 61273304, 61202170) and the Research Fund for the Doctoral Program of Higher Education of China (No. 20130072130004).

References

[1] G. Birkhoff, Lattice Theory, Colloquium Publication, American Math Society, Providence, R.I., 1940.
[2] A. Burusco, R. Fuentes-Gonzales, Construction of the l-fuzzy concept lattice, Fuzzy Sets Syst. 97 (1998) 109–114.
[3] J.K. Chen, J.J. Li, Y.J. Lin, G.P. Lin, Z.M. Ma, Relations of reduction between covering generalized rough sets and concept lattices, Inf. Sci. 304 (2015) 16–27.
[4] S.M. Dias, B.M. Nogueira, L.E. Zarate, Adaptation of FCANN method to extract and represent comprehensible knowledge from neural networks, Stud. Comput. Intell. 134 (2008) 163–172.
[5] B. Ganter, R. Wille, Formal Concept Analysis:Mathematical Foundations, Springer-Verlag, Berlin, 1999.
[6] Q.H. Hu, Z.X. Xie, D.R. Yu, Hybrid attributes reduction based on a novel fuzzy-rough model and information granulation, Pattern Recog. 40 (2007) 3509–3521.
[7] L. Jiang, J. Deogun, SPICE: a new frame work for data mining based on probability logic and formal concept analysis, Fundamenta Informaticae 78 (4) (2007) 467–485.
[8] X.P. Kang, D.Y. Li, S.G. Wang, K.S. Qu, Formal concept analysis based on fuzzy granularity base for different granulations, Fuzzy Sets Syst. 203 (2012) 33–48.
[9] X.P. Kang, D.Y. Li, S.G. Wang, K.S. Qu, Rough set model based on formal concept analysis, Inf. Sci. 222 (2013) 611–625.
[10] X.P. Kang, D.Y. Li, S.G. Wang, Research on domain ontology in different granulations based on concept lattice, Knowl. Based Syst. 27 (2012) 152–161.
[11] R.E. Kent, Rough concept analysis, Fundamenta Informaticae 27 (1996) 169–181.
[12] H.L. Lai, D.X. Zhang, Concept lattices of fuzzy contexts: formal concept analysis vs. rough set theory, Int. J. Approx. Reason. 50 (5) (2009) 695–707.
[13] H.S. Lee, An optimal algorithm for computing the max-min transitive closure of a fuzzy similarity matrix, Fuzzy Sets Syst. 23 (2001) 129–136.
[14] J.H. Li, Y. Ren, C.L. Mei, Y.H. Qian, X.B. Yang, A comparative study of multigranulation rough sets and concept lattices via rule acquisition, Knowl. Based Syst. 91 (2016) 152–164.
[15] J.H. Li, C.L. Mei, L.D. Wang, J.H. Wang, On inference rules in decision formal contexts, Int. J. Comput. Intell. Syst. 8 (1) (2015a) 175–186.
[16] J.H. Li, C.L. Mei, W.H. Xu, Y.H. Qian, Concept learning via granular computing: a cognitive viewpoint, Inf. Sci. 298 (2015b) 447–467.
[17] J.H. Li, C.L. Mei, J.H. Wang, X. Zhang, Rule-preserved object compression in formal decision contexts using concept lattices, Knowl. Based Syst. 71 (2014) 435–445.
[18] T.Y. Lin, Neighborhood systems and relational database, in: Proceedings of the 16th ACM Annual Conference on Computer Science, Altanta, 1988, pp. 725–728.
[19] Y. Leung, D.Y. Li, Maximal consistent block techniques for rule acquisition in incomplete information systems, Inf. Sci. 153 (2003) 85–106.
[20] J.Y. Liang, Z.Z. Shi, D.Y. Li, M.J. Wierman, Information entropy, rough entropy and knowledge granulation in incomplete information system, Int. J. Gen. Syst. 35 (6) (2006) 641–654.
[21] M. Liu, M.W. Shao, W.X. Zhang, C. Wu, Reduct method for concept lattices based on rough set theory and its application, Comput. Math. Appl. 53 (9) (2007) 1390–1410.
[22] J.M. Ma, Y. Leung, W.X. Zhang, Attribute reductions in object-oriented concept lattices, Int. J. Mach. Learn. Cybern. 5 (2014) 789–813.
[23] J.S. Mi, Y. Leung, W.Z. Wu, Approaches to attribute reduct in concept lattices induced by axialities, Knowl. Based Syst. 23 (6) (2010) 504–511.
[24] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Dordrecht, 1991.
[25] J. Pócs, Note on generating fuzzy concept lattices via galois connections, Inf. Sci. 185 (1) (2012) 128–136.
[26] J.J. Qi, T. Qian, L. Wei, The connections between three-way and classical concept lattices, Knowl. Based Syst. 91 (2016) 143–151.
[27] Y.H. Qian, J.Y. Liang, C.Y. Dang, Incomplete multi-granulation rough set, IEEE Trans. Syst. Man Cybern. Part A 40 (2) (2010) 420–431.
[28] A.K. Sarmah, S.M. Hazarika, S.K. Sinha, Formal concept analysis: current trends and directions, Artif. Intell. Rev. 44 (1) (2015) 47–86.
[29] M.W. Shao, Y. Leung, Relations between granular reduct and dominance reduct in formal contexts, Knowl. Based Syst. 65 (2014) 1–11.
[30] H. Thiele, On semantic models for investigating computing with words, in: Second International Conference on Knowledge Based Intelligent Electronic Systems, Adelaide, 1998, pp. 32–98.
[31] A.H. Tan, J.J. Li, G.P. Lin, Connections between covering-based rough sets and concept lattices, Int. J. Approx. Reason. 56 (2015) 43–58.
[32] V. Ventos, H. Soldano, Alpha galois lattices: an over view, in: Proceedings of ICFCA, 2005, pp. 299–314.
[33] Q. Wan, L. Wei, Approximate concepts acquisition based on formal contexts, Knowl. Based Syst. 75 (2015) 78–86.
[34] L.D. Wang, X.D. Liu, Concept analysis via rough set and AFS algebra, Inf. Sci. 178 (21) (2008) 4125–4137.
[35] L. Wei, J.J. Qi, Relation between concept lattice reduct and rough set reduct, Knowl. Based Syst. 23 (8) (2010) 934–938.
[36] R. Wille, in: I. Rival (Ed.), Restructuring lattice theory: an approach based on hierarchies of concepts, Ordered Sets, Dordrecht, Reidel, 1982, pp. 445–470.
[37] Q. Wu, Z.T. Liu, Real formal concept analysis based on grey-rough set theory, Knowl. Based Syst. 22 (1) (2009) 38–45.

[38] W.Z. Wu, Y. Leung, J.S. Mi, Granular computing and knowledge reduction in formal contexts, IEEE Trans. Knowl. Data Eng. 21 (10) (2009) 1461–1474.
[39] Y.Y. Yao, Information granulation and rough set approximation, Int. J. Intell. Syst. 16 (1) (2001) 87–104.
[40] L.A. Zadeh, Fuzzy sets and information granularity, in: N. Gupta, R. Ragade, R. Yager (Eds.), Advances in Fuzzy Set Theory and Application. Amsterdam: North-Holland, 1979, pp. 3–18.
[41] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets Syst. 90 (2) (1997) 111–127.
[42] Y.H. Zhai, D.Y. Li, K.S. Qu, Fuzzy decision implications, Knowl. Based Syst. 37 (2013) 230–236.
[43] Y.H. Zhai, D.Y. Li, K.S. Qu, Decision implication canonical basis: a logical perspective, J. Comput. System Sci. 81 (2015) 208–218.
[44] B. Zhang, L. Zhang, Theory and applications of problem solving, North-Holland: Elsevier Science Publishers B.V., 1992.