Knowledge-Based Systems 102 (2016) 103–115

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A variable precision rough set model based on the granularity of tolerance relation

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 10 December 2015
Revised 25 February 2016
Accepted 30 March 2016
Available online 1 April 2016

Keywords:
Rough set
Concept lattice
Tolerance relation
Tolerance class

ABSTRACT

As one of core problems in rough set theory, normally, classification analysis requires that “all” rather than “most”elements in one class are similar to each other. Nevertheless, the situation is just opposite to that in many actual applications. This means users actually just require “most” rather than “all”elements in a class are similar to each other. In the case, to further enhance the robustness and generalization ability of rough set based on tolerance relation, this paper, with concept lattice as theoretical foundation, presents a variable precision rough set model based on the granularity of tolerance relation, in which users can flexibly adjust parameters so as to meet the actual needs. The so-called relation granularity means that the tolerance relation can be decomposed into several strongly connected sub-relations and several weakly connected sub-relations. In essence, classes defined by people usually correspond to strongly connected sub-relations, but classes defined in the paper always correspond to weakly connected sub-relations. In the paper, an algebraic structure can be inferred from an information system, which can organize all hidden covers or partitions in the form of lattice structure. In addition, solutions to the problems are studied, such as reduction, core and dependency. In short, the paper offers a new idea for the expansion of classical rough set models from the perspective of concept lattice.

1. Introduction

It is known that concept has been taken as the unit or cell of human cognition in people’s thinking activities, since concept contains the most essential information of some kind of things, it plays an important role in human’s cognitive process. In essence, as one major method for human to know the real world and its laws, concept thinking can be served as the foundation for people to form various complicated ideas and also effective means to express knowledge. In 1982, German mathematician Wille professor brought forth formal concept analysis (FCA), or concept lattice theory [46], which can be considered as an application branch of lattice theory. As one kind of method to mathematically abstract and formalize concepts from the objective world, FCA greatly stimulates people’s enthusiasm to solve problems under the concept thinking. In FCA, the basic viewpoint of concept essentially developed from the understanding of concept in philosophy, that is, one concept is mathematically described from aspects of extent and intent, in which extent refers to the set of objects covered by concept, and intent refers to the set of common characteristics of objects covered by concept. Concept lattice, as the core data structure of FCA, is an effective tool for data analysis and rule extraction, and can vividly and concisely manifest the generalization-specialization relationship among concepts by means of Hasse graph. In recent years, concept lattice has developed into a powerful data analysis method [14,16,21,25,32,37,48,52], and found wide applications in many fields like data mining analysis, information retrieval, knowledge discovery, ontology engineering, etc.

In practice, information collected from actual systems often contains noise, namely, information is not always accurate or complete. Along with the rapid development of science and technology, the uncertainty of information is more and more remarkable. Therefore, it is always inevitable for people to process the uncertainty and incomplete information in various applications. In the case, how to distill useful knowledge from the massive, inaccurate, fuzzy or incomplete information has become an extremely urgent task. Although, people can use pure mathematical assumptions to eliminate or avoid this uncertainty, but the effect is often not ideal. Conversely, if methods can appropriate to deal with these information, it is often helpful to solve many complex practical problems. Over the years, researchers have been trying to find effective ways to deal with the incomplete and uncertainty information

* Corresponding author. Tel.: +86 0351 6563220.
E-mail addresses: tongji_kangxp@sina.com (X. Kang), miaoduoqian@163.com (D. Miao).

http://dx.doi.org/10.1016/j.knosys.2016.03.030

scientifically. As classic methods to deal with uncertain information, evidence theory, fuzzy theory, probability statistics, etc. have been used in many practical fields. However, these methods need some additional information or prior knowledge, such as fuzzy membership function, belief function, statistical distribution function, etc. which can not be easily obtained.

In 1982, Polish scholar Pawlak brought forth rough set theory [31], as a kind of important reasoning technology in artificial intelligence, which can effectively analysis and process the fuzzy and uncertain information without any prior knowledge except for data sets. Its main idea is to, with the classification ability being kept unchanged, deduce decision or classification rules of problems through knowledge reduction. Meanwhile, it can use the observed and measured knowledge to approximately describe imprecise or uncertain concepts. Due to its effectiveness and usability in the process of dealing with uncertain problems, rough set has already drawn much attention of scholars [9,11,12,33], and lots of research results have been widely applied to various fields, such as medical diagnosis, decision analysis, image processing, machine learning, and so on. In addition, with the deepening research and widening scope, the data forms and organization structures are increasingly diversified, so it becomes more and more difficult for people to effectively solve the complicated practical problems just through any single theory. Therefore, combining rough set with other artificial intelligence technology has become a hot research topic of international scholars, such as probability statistics, fuzzy set, evidence theory, neural network, concept lattice [35,39,44,47], and so on. So far, the whole theoretical system of rough set has already been gradually maturing and increasingly perfect, which greatly enriched and expanded the theoretical foundation and the application scope of rough set.

Rough set and concept lattice, as two mathematical branches generated in the same era, there are some significant differences from the perspective of their research methodology, but the same research background and objective indicate that they must have something in common. In fact, the two theories share many similarities [17], such as any one-valued formal context, as a kind of data set, is just a special case of information systems essentially, therefore, their mutual reference and integration not only enhance their own analytic abilities, but also can help to understand one theory from the perspective of another. Meanwhile, by means of mixing their respective advantages, the fusion theory may help to establish a more general and universal data analysis framework. Therefore, it is extremely significant to combine two theories in terms of their advantages. Recently, many remarkable research achievements of the fusion theory have been made. Oosthuizen informally described the connection between rough set and concept lattice [30]. In the study of logical models, Duntsch and Gediga defined modal-style operators on the basis of binary relations, and constructed the attribute-oriented concept lattice according to the upper approximate operator [5,8]. Deogun and Saquer mainly discussed the monotone concept lattice, which is a direct expansion of classical concept lattice [4,36]. By introducing the idea of upper and lower approximations in rough set, Yao expended the definition of concept lattice, studied the rough set approximation of formal concept, built object-oriented and attribute-oriented concept lattices, and proved that attribute-oriented concept lattice and object-oriented concept lattice are isomorphic [49–51]. Zhang et al. introduced variable threshold concept lattices [53]. Belohlavek et al. provided the uniform structure of different variable threshold concept lattices [1]; Fan et al. studied fuzzy inferences based on fuzzy concept lattices [6]. Through comparing the relationship between fuzzy concept lattice and rough set, Lai et al. pointed out that each complete fuzzy concept lattice could be expressed as the concept lattice in the sense of rough set under certain conditions [18]. Lots of scholars introduced the idea of reduction in rough set into concept lattice, and discussed the reduction theory in concept lattice [2,20,24,27–29,45]. Kang et al. once suggested a rough set model based on concept lattice, which solved the problem of algebraic structure in the discrete information system, namely inducing a lattice structure from an information system, with each node in the lattice being called a rough concept, meanwhile, they also presented solutions to some common problems in rough set based on concept intents, such as core, reduction and function dependence [15]. For more flexible and efficient learning concept, from the cognitive computing perspective, Li et al. investigated concept learning by means of granular computing and set approximations [22], in addition, they have focused on issues of approximate concept lattice, approximate decision rule and knowledge reduction in incomplete decision contexts [23]. Shao and Leung revealed some relationships of reduction results in rough set and concept lattice [38]. Tan systematically explored connections between rough set and concept lattice in terms of approximation operators, structures and knowledge reduction [43]. Li et al. [26] made a comparison between multigranulation rough sets and concept lattices via rule acquisition, and obtained some interesting results. For more research findings concerning the fusion theory of rough set and concept lattice, please see the literature [51].

It is known to us that Pawlak's classical rough set model is established on the basis of equivalence relation (equivalence relation needs to meet reflexivity, transitivity and symmetry), and used to process complete information systems containing nominal attributes (domain of attribute is composed of several discrete values, and different values are independent of each other). However, when the domain of attribute is a real number set, or the differences among different values are caused by test errors, or the problems to be solved are highly complicated, or the scale of data set is too big, it is meaningless to analyze some minor differences. In the case, classical Pawlak's rough set model obviously has some limitations. In practical applications, users may not only require that objects with identical attribute values should be put into the same class, but also assume that objects with similar attribute values should also be classified the same. To further enhance the data processing capability of rough set, many scholars expand equivalence relations to tolerance relations (sometimes called similarity relation) only meeting reflexivity and symmetry. Tolerance relation is substantially different from binary relation of other types in terms of symmetry. Namely, symmetry is the basic characteristic of tolerance relation. In view of the universality of tolerance relation, great research findings have been made on the theory and application of rough set based on tolerance relation in recent years. To enhance the data processing capability of rough set, Slowinski et al. studied the properties and applications of rough set based on similarity relation, and pointed out that rough set based on similarity relation can be used for ignoring minor differences of attribute values [41,42]; Skowron et al. presented rough set based on tolerance relation, which was conducive to enhancing the robustness of system decisions and also the efficiency of decision making [40]; Leung and Li [19] studied the granules in incomplete information system, namely, with maximal tolerance classes as information granules, overcame the flaws of knowledge expression based on similarity class. Hu et al. proposed neighborhood rough set models in information systems with mixed features, where objects with numerical attributes were granulated with fuzzy tolerance relations obtain by Euclidean distance, while objects with nominal features were granulated with equivalence relations [13]. Guan and Wang applied maximal tolerance classes to set-valued information system, and discussed problems of attribute reduction and decision rule acquisition [10]. Based on maximal tolerance classes, Qian et al. studied the approximation reduction in inconsistent incomplete decision tables [34]. Dai defined fuzzy tolerance

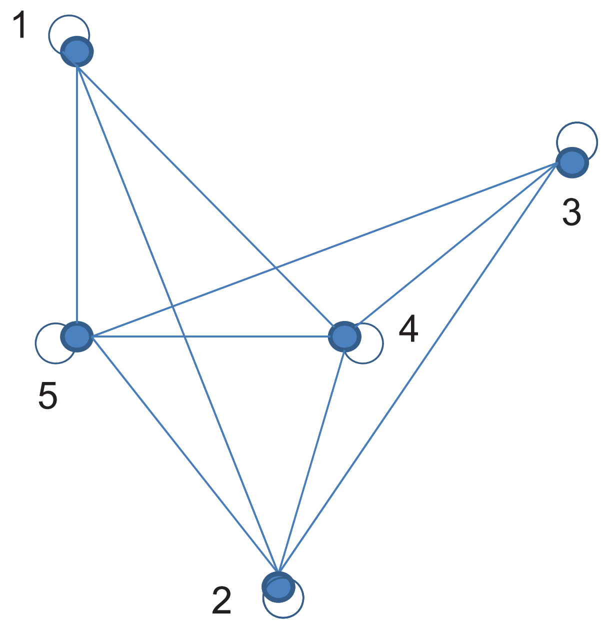

Fig. 1. A tolerance relation graph.

relations in incomplete numerical data, established the fuzzy tolerance rough set, and discussed the problem of attribute reduction [3].

Constructing and discovering new classification models has became an effective mean for expanding rough set, there have been lots of related research results in recent years. Among various kinds of models, the most common one is defined as: let R be a tolerance relation on U, then

"$\mathscr{E}$ is a class, if and only if $\mathscr{E}$ is a maximal set satisfying $\mathscr{E} \times \mathscr{E} \subseteq R$."

In another way, above type of classes essentially can be understood from the perspective of relation granularity. Namely, the tolerance relation R can be granulated into several sub-relations defined as follows:

Definition 1. Let $H \times N \subseteq R$, if $\nexists H_1 \times N_1$ with $H_1 \times N_1 \subseteq R$ and $H \times N \subset H_1 \times N_1$, then $H \times N$ is called a sub-relation of R. For any sub-relation $H \times N$, if $H = N$, then $H \times N$ is called a strongly connected sub-relation; if $H \neq N$, then $H \times N$ is called a weakly connected sub-relation.

Apparently, the definition mentioned-above shows R can be granulated into several strongly connected sub-relations and weakly connected sub-relations. For instance, Fig. 1 is a tolerance relation graph, in which the sub-relations 1245 × 1245 and 2345 × 2345 are strongly connected, the sub-relation 245 × 12345 is weakly connected (here, $\{p_1, p_2, \ldots, p_n\}$ is simply as $p_1 p_2 \ldots p_n$). Furthermore, from above definition, there exists following conclusion

- $\mathscr{E}$ is a class, if and only if $\mathscr{E} \times \mathscr{E}$ is a strongly connected sub-relation.

If problems are too complicated or solving them requires high cost or data sets have some noise, people may no longer focus on some detailed information and take exact solutions as the goal, but replace exact solutions with the feasible approximate solutions, this will surely help to simplify problems and enhance efficiency.

As one of central problems of rough set, classification problems usually require that "all" rather than "most" elements in a class are similar to each other. Obviously, when people explores approximate solutions rather than accurate ones in some problems, the condition for generating classical classes is too restrictive to some extent, this may result in too small classical classes, which perhaps seriously affect the generalization ability of classification algorithms of rough set. In the case, users possibly just require that "most" rather than "all" elements in a class are similar to each other. Therefore, to further enhance the robustness and generalization ability of rough set based on tolerance relation, the classical class mentioned-above is further extended to the θ-class defined in the paper. That is, a θ-class $\mathscr{C}$ needs to meet the following condition.

- there exists maximal subset $L \subseteq \mathscr{C}$ satisfying

  $$L \times \mathscr{C} \subseteq R \text{ and } |\mathscr{C}| \times \theta \leqslant |L|$$

  where R is a tolerance relation on U, $0 \leq \theta \leq 1$. In the case, we say L is the class core of $\mathscr{C}$.

Above definition just lists the condition that θ-class must to meet, and it is only the summary description of θ-class. For how to construct θ-classes, the paper will give a detailed description in the following chapters.

In above definition, the description of θ-classes not requires that "all" elements in one class are similar to each other. It just requires that "most" elements in a class are similar to each other, namely, it only requires $L \times \mathscr{C} \subseteq R$ rather than $\mathscr{C} \times \mathscr{C} \subseteq R$. Obviously, there is following fact

- for any classical class $\mathscr{E}$, there always exists a θ-class $\mathscr{C}$ with $\mathscr{E} \subseteq \mathscr{C}$.

In Fig. 1, the classical classes {1, 2, 4, 5} and {2, 3, 4, 5} are shown in Fig. 2(a), and the θ-class with $\theta = 0.6$ is shown in Fig. 2(b).

In fact, classical classes defined by people essentially correspond to strongly connected sub-relations, but θ-classes constructed in the paper always correspond to weakly connected sub-relations. Based on the points discussed above, an algebraic structure can be further inferred from an information system, which can organize all hidden covers or partitions in the form of lattice structure. Furthermore, solutions to the problems are studied, such as reduction, core and dependency. In short, the method presented in the paper offers another way for expanding rough set, meanwhile, it also helps to reasonably analyze and interpret rough set from the perspective of concept lattice.

The following chapters are arranged as follows: Section 2 gives a brief introduction on concept lattice and rough set; Section 3 introduces parameters, studies the granularity of tolerance relations, expands classical classes to θ-classes, and derives an one-valued formal context from an information system; Section 4 builds an algebraic structure from an information system based on concept lattice; Section 5 mainly discusses methods to deal with common problems from the perspective of concept lattice, such as attribute reduction, core and dependency; Section 6 is a brief conclusion and outlook.

2. Concept lattice and rough set

This section briefly introduces concept lattice and rough set, about more detailed information, please refer to [7,31].

A formal context is a triple (G, M, I), where G is a finite nonempty set of objects, M is a finite nonempty set of attributes, I is a binary relation between G and M, namely $I \subseteq G \times M$.

Definition 2. In $K = (G, M, I)$, the map $\mathscr{P}(G) \to \mathscr{P}(M)$ is defined as: let $A \in \mathscr{P}(G)$, then

$$A' = \{m \in M \mid (g, m) \in I, \forall g \in A\}$$

the map $\mathscr{P}(M) \to \mathscr{P}(G)$ is defined as: let $B \in \mathscr{P}(M)$, then

$$B' = \{g \in G \mid (g, m) \in I, \forall m \in B\}$$

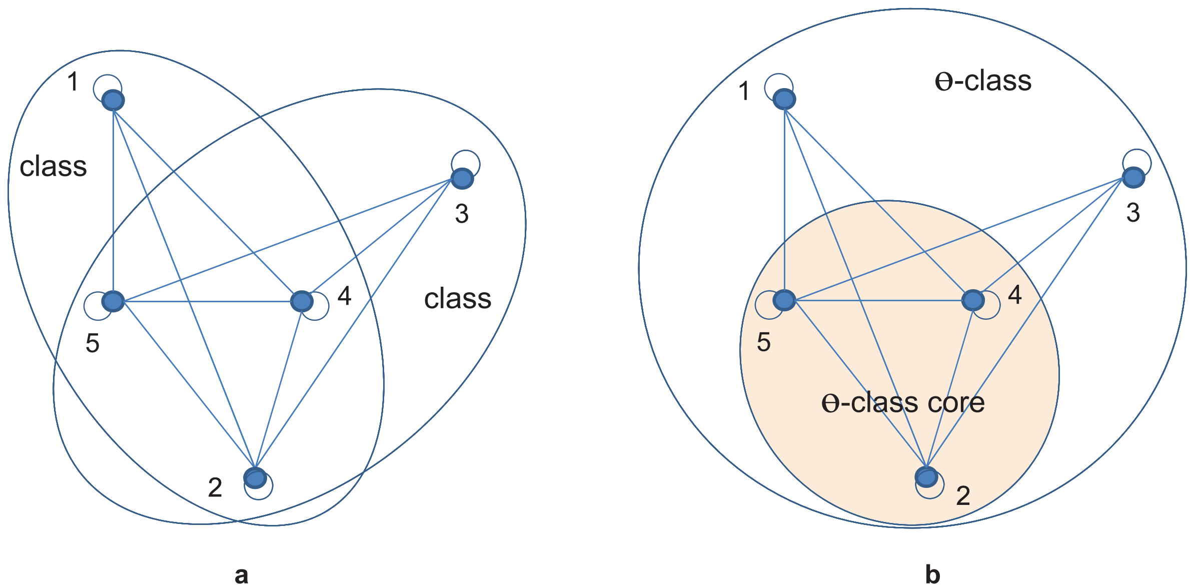

Fig. 2. Classical classes and θ-class derived from Fig. 1 with $\theta = 0.6$

where $\mathscr{P}(G)$ and $\mathscr{P}(M)$ be power sets of G and M respectively.

In above definition, if $A' = B$ and $B' = A$, then we say (A, B) is a formal concept. The order relationship "$\preceq$" between concepts (A_1, B_1) and (A_2, B_2) is described as

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

In essence, $(\mathscr{B}(K), \preccurlyeq)$ is a complete lattice, where $\mathscr{B}(K)$ is the set of all concepts in K. Meanwhile, there exist following conclusions as follows

Proposition 1. *Let (G, M, I) be a formal context, $A, A_1, A_2 \subseteq G$, $B, B_1, B_2 \subseteq M$, 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. *Let $K = (G, M, I)$ be a formal context, $B \subseteq M$, then*

$$B'' = \cap\{T \in \mathscr{U} | B \subseteq T\}$$

where $\mathscr{U}$ is the set of intents of all concepts in K.

Normally, an information system is formalized as a quadruple (U, AT, V, f), where U called universe is a finite nonempty set of objects, AT is a finite nonempty set of attributes, $V = \bigcup_{m \in AT} V_m$ with V_m called the domain of attribute m, f: $U \times AT \to V$ is a function, that is, $f(x, m) \in V_m$ for any $x \in U, m \in AT$. In the following, (U, AT, V, f) is simplified as (U, AT).

In an information system (U, AT), there always exists a binary relation R_B for any $B \subseteq AT$. In this case, let $m \in B$, if $R_B \neq R_{B-m}$, then we say m is indispensable in B, and $core(B)$ consisted of all indispensable attributes in B is called the core of B; let $B, D \subseteq AT$, $B \subseteq D$, if B is a minimum-subset satisfying $R_B = R_D$, then we say B is a reduction of D; let $B, C \subseteq AT$, if $R_B \subseteq R_C$, then we say $B \to C$ is a dependency.

3. One-valued formal contexts derived from information systems

It is known to us that Pawlak's classical rough set is established on the basis of equivalence relation and used to process only nominal attributes. However, some common data sets always contain numeric attributes (the domain of attribute is a real number set or a subset of the real number set) in actual applications. In the case, the classical rough set obviously has some limitations. Specially, for any numeric attribute, it is rare that different objects possess completely identical attribute values. And further, too small equivalence classes formed through equivalence relation would seriously affect the generalization ability of classification algorithms. To address above problem, many scholars take information systems containing numeric attributes as research background, and expand equivalence relations to tolerance relations.

Table 1
A typical information system.

	a	b	c	d
1	yes	u_1	v_1	yes
2	yes	u_2	v_2	yes
3	yes	u_3	v_3	no
4	yes	u_4	v_4	no
5	no	u_5	v_5	no
6	no	u_6	v_6	no
7	no	u_7	v_7	yes
8	no	u_8	v_8	yes

Table 1 is a typical information system, where a and d are nominal attributes, b and c are numeric attributes. In fact, for any numeric attribute m, there are great research findings have been made to obtain the tolerance relation I_m on V_m. For example, a simple method is described as

Let $|V_m| = n$, for any v_i, $v_j \in V_m$, if $sim(v_i, v_j) \geq \delta_m$, then $(v_i, v_j) \in I_m$, where

$$sim(v_i, v_j) = 1 - \frac{|v_i - v_j|}{\max\{v_1, v_2, \ldots, v_n\}}, \quad 0 \leqslant \delta_m \leqslant 1$$

Since the solution procedure for generating tolerance relations is not the focus of the paper, we will not elaborate on how the tolerance relation I_m on V_m is obtained, that is, I_m is directly given in the paper.

3.1. Concept scales and concept scaling

Normally, the classical concept lattice is always used to study one-valued formal contexts, but information systems are many-valued formal contexts essentially. Therefore, to study information

Table 2
Tolerance scales S_b and S_c relative to b and c separately.

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8		v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
u_1	×		×	×					v_1	×		×	×	×	×		
u_2		×	×	×					v_2		×		×	×	×		
u_3	×	×	×	×	×				v_3	×		×	×	×	×		
u_4	×	×	×	×	×	×			v_4	×	×	×	×	×	×		
u_5			×	×	×	×			v_5	×	×	×	×	×	×	×	×
u_6				×	×	×			v_6	×	×	×	×	×	×	×	
u_7							×	×	v_7					×	×	×	×
u_8							×	×	v_8					×		×	×

Table 3
The equivalence scale S_a relative to a.

	no	yes
no	×	
yes		×

Table 4
A concept scale of an attribute without any prior knowledge.

	low	high	very high
low	×		
high		×	
very high			×

Table 5
A concept scale of an attribute with some prior knowledge.

	low	high	very high
low	×		
high		×	
very high		×	×

systems with operators in Definition 2, we should transform information systems into one-valued formal contexts. Recently, there have been many ways of transforming an information system into an one-valued context, such as concept scaling technology and logic scaling technology. In the paper, we select the former.

The basic idea of concept scaling is to transform an information system into an one-valued formal context via concept scales. Each attribute corresponds to one concept scale, and different attributes may correspond to different concept scales. The so-called concept scale of an attribute can be viewed as the interpretation or detailed description of the domain of attribute. That is, for any attribute m, the corresponding concept scale essentially embodies the relationship among different values in V_m. Concept scaling can be simply understood as the strategy for transforming an information system into an one-valued formal context on the basis of these concept scales. The definition of concept scales and the selection of concept scaling strategies directly determine the one-valued formal context transformed. In fact, in the whole process of transformation, concept scales only play the intermediary role, rather the final derivative context.

Definition 3. In $S=(U,AT)$, we say $S_m=(G_m,M_m,I_m)$ is a concept scale of $m \in AT$, if $V_m \subseteq G_m$.

The definition above only requires $V_m \subseteq G_m$, without any more restrictions on G_m and M_m. In order to meet some actual situations, other reasonable restrictions should be considered as well to construct S_m.

In Definition 3, if $G_m = M_m = V_m$, then S_m is called a relation-concept scale. Let S_m be a relation-concept scale, if I_m is an equivalence relation on V_m, then we say S_m is an equivalence scale; if I_m is a tolerance relation on V_m, then S_m is called a tolerance scale. For instance, for the numeric attributes b and c in Table 1, the corresponding tolerance scales S_b and S_c are shown in Table 2; for the nominal attribute a (or d), the corresponding equivalence scale S_a(or S_d) is shown in Table 3. In above tables, if the crossing of $v \in V_m$ row and $w \in V_m$ column is denoted as " × ", then it means $(v, w) \in I_m$.

In fact, there are many different ways of concept scaling, this means that we can flexibly carry out concept scaling according to characteristics of information systems and the corresponding domain knowledge. In the paper, a relatively simple method for concept scaling (called for short as plain scaling) is adopted, which is formalized as follows.

In $S=(U,AT)$, let $S_m=(V_m,V_m,I_m)$ be the relation-concept scale of attribute m, then for any $(x, y) \in U \times U$ and $m \in AT$, by the transformation rule

$$((x,y),m) \in J \Leftrightarrow (w,v) \in I_m,\ \ f(x,m)=w,\ f(y,m)=v$$

S can be converted to an one-valued formal context

$$K_S=(U \times U, AT, J)$$

K_S is called the derivative context of S. In the paper, from the whole transformation process shown in Fig. 3, we can see that the concept scales and concept scaling play significant roles.

In the above transformation, for any values v, $w \in V_m$, the concept scale can filter out there own information, but care about whether they meet $(v, w) \in I_m$. Namely, S_m is essentially an one-valued formal context, which only remains the relationship among different values of V_m in the form of one-valued formal context.

It is worth to mention that concept scale is not only a simple transformation tool, may also contribute to solve actual problems more scientifically and rationally. It is known that classical rough set always does not require additional prior knowledge during the procedure of data analysis and processing, this is an advantage, but is also its shortcoming, that is, when dealing with data that requires additional information or knowledge, classical rough set will be helpless. For instance, in classical rough set, for an attribute "blood pressure", the corresponding attribute values "low", "high" and "very high" are normally independent of one another by default, namely, there is no relationship between above attribute values. In the case, the corresponding concept scale is shown in Table 4. However, there may exist relationship between "high" and "very high" in many applications, namely, if one patient's blood pressure is "very high", then it must be "high". In the case, the corresponding concept scale is shown in Table 5. Obviously, by means of concept scales, we can introduce some prior knowledge to the final derivative context, rather than the simple transformation, and then the actual problems can be solved more reasonably based on concept lattice. The related content mentioned-above will be one of the focus of our next research.

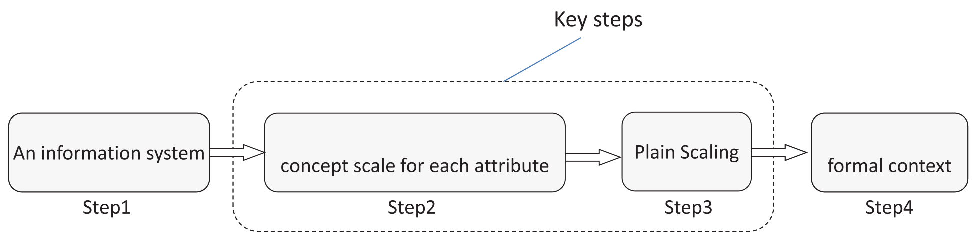

Fig. 3. The process of one-valued contexts derived from an information system.

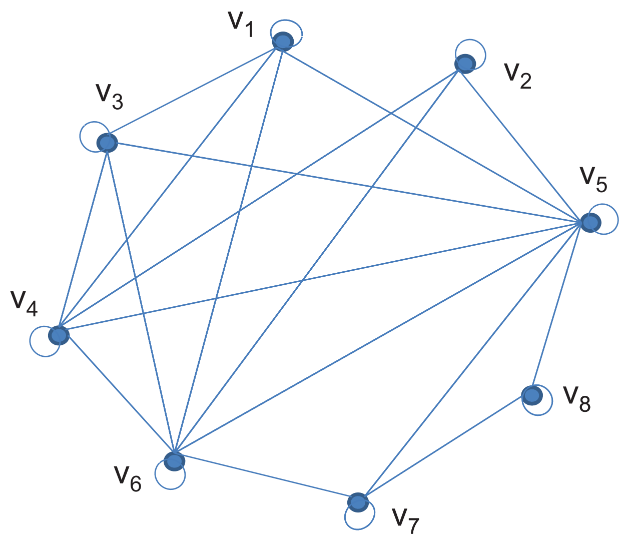

Fig. 4. The undirected graph derived from Table 2(b).

3.2. Granularity of tolerance relations and variable precision concept scales

In fact, we can view the tolerance scale $S_m = (V_m, V_m, I_m)$ as a relation graph, where V_m denotes the set of all vertexes, and I_m is the set of all undirected edges. For instance, S_c shown in Table 2 (b) can be interpreted as the undirected graph shown in Fig. 4. Clearly, in the worst case, if $|V_m| = n$, then $|I_m| = n \times (n-1)/2$. It is obvious that a bigger n indicates massive edges, and then eventually resulted in a huge and complicated graph. This not only brings about higher time and space complexity for further computing and solving problems, but also does not help to further understanding the schematic configuration of graph on macroscopic view. In the case, to better understand and solve problems rather than get lost in unnecessary details, the paper presents a new method based on the granularity of V_m, in which the granularity of I_m plays an important role.

When processing large scope and complicated information, people, due to their limited cognitive capability, usually divide the complicated information into several simple blocks in terms of its characteristics and performance so as to better analyze and solve actual problems. Each block is always considered as a granule (or a class). In fact, the granularity mechanism is also one major feature of human thinking. As one core problems of rough set, the information granularity always serves as the foundation of knowledge acquisition, so discovering and constructing new granularity models is an effective mean for xpanding rough set theory. For instance, the complicated tolerance relation graph shown in Fig. 4 can be granulated into several relatively simple sub-graphs shown in Fig. 5, and any sub-graph can be viewed as a class. For any sub-graph, we can see there always exists a undirected edge between any two vertexes.

Normally, classification analysis requires that “all” rather than “most” elements in one class are similar to each other. Nevertheless, in many actual applications, users possibly just require that “most” rather than “all” elements in one class are similar to each other. Therefore, to further enhance the robustness and generalization ability of rough set, this paper, with concept lattice as the theoretical foundation, discusses the classification problem in V_m by means of the granularity of I_m, and further expands classical classes into θ-classes. It is worth mentioning that the θ-classes referred here means the classes in V_m rather than the ones in U. Hereinafter, by referring to Definition 1, the tolerance relation I_m can be further decomposed into several strongly connected sub-relations and several weakly connected sub-relations. In essence, classical classes defined by people usually correspond to strongly connected sub-relations, but classes defined in the paper always correspond to weakly connected sub-relations.

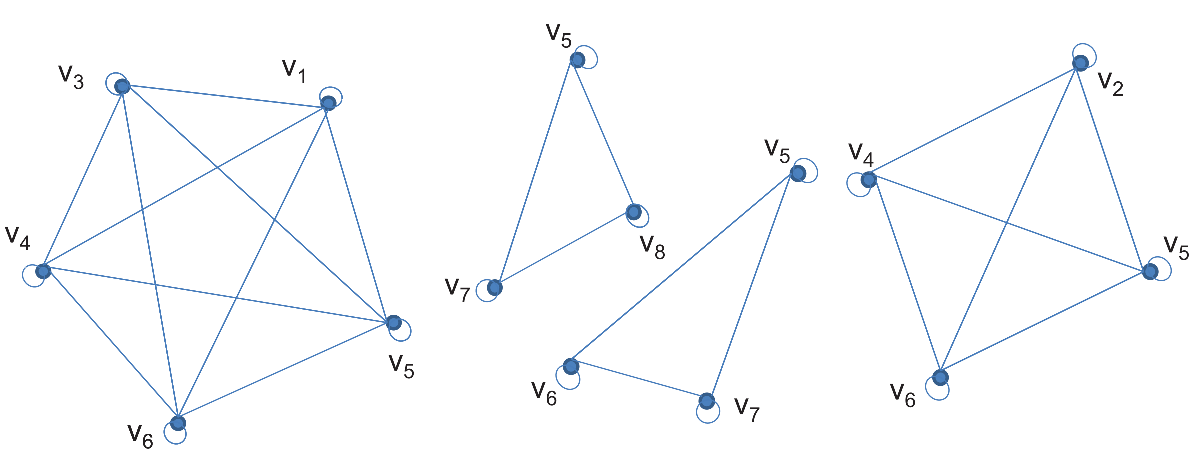

Fig. 5. Some sub-graphs derived from Fig. 4.

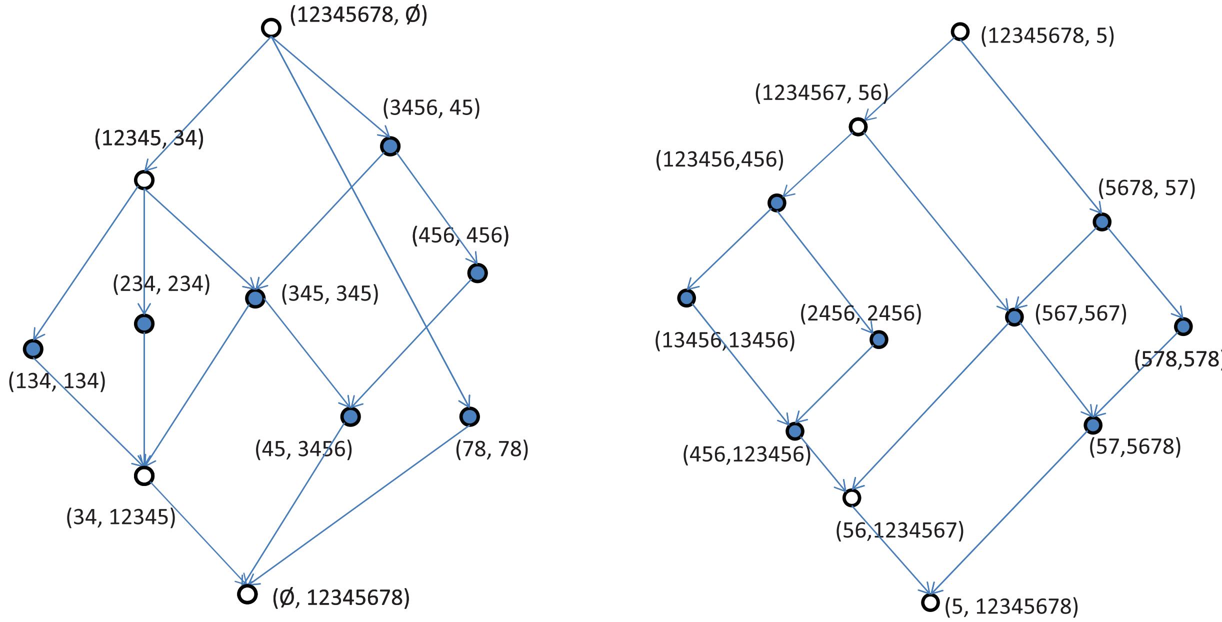

Fig. 6. $\mathscr{B}(S_b)$ and $\mathscr{B}(S_c)$ derived from Table 2.

Let $\Sigma(I_m)$ be the set of all sub-relations in I_m, then the following conclusion can be inferred easily

Theorem 1. *Let $S_m = (V_m, V_m, I_m)$ be a tolerance scale, H, $N \subseteq V_m$, then*

$$(H, N) \in \mathscr{B}(S_m) \Leftrightarrow H \times N \in \Sigma(I_m)$$

Proof. The conclusion can be deduced from Definition 1 and Definition 2 immediately. □

Apparently, the theorem above shows that $\Sigma(I_m)$ can obtained by calculating $\mathscr{B}(S_m)$. Taking scales in Table 2 as examples, by means of operators in Definition 2, corresponding concept lattice structures can be calculated easily, which are shown in Fig. 6 (for convenience, any v_i or u_i is simplified as "i").

Definition 4. In $S_m = (V_m, V_m, I_m)$, let $0 \le \alpha \le 1$, for any $H \times N \in \Sigma(I_m)$, the connected-degree between H and N is defined as

$$\phi_\alpha(H \times N) = \frac{|H \cap N|}{|H \cap N| + \alpha \times |(H - N) \cup (N - H)|}$$

In above definition, $H - N$ denotes the set of elements appearing in H but not in N; $N - H$ denotes the set of elements appearing in N but not in H. In fact, $(H - N) \cup (N - H)$ is the different part between H and N, and $H \cap N$ is the common part between H and N. In addition, the parameter α, to a given degree, reflects that the influence of the common part on the connected-degree is bigger than the different part. Obviously, the bigger common part is, the connected-degree is; on the contrary, the smaller common part is, the lower connected-degree is.

Definition 5. In $S_m = (V_m, V_m, I_m)$, let $0 \le \theta \le 1$, if $\mathscr{O} = H \cup N$ is the maximum-subset satisfying

$H \cup N \in \Sigma(I_m)$ and $\phi_\alpha(H \times N) \ge \theta$

then we say $\mathscr{O}$ is a θ-class. In the case, we say $H \cap N$ is the class core of $\mathscr{O}$.

In above definition, when $S_m = (V_m, V_m, I_m)$ is viewed as a tolerance relation graph, then the θ-class $\mathscr{O}$ can be essentially interpreted as the sub-graph (H, N, $H \times N$) in S_m. In the case, $\phi_\alpha(H \times N)$ can be understood as quantitative measurement for the scale of undirected edges in the sub-graph $\mathscr{O}$ essentially. Since it is of no practical significance to view $\mathscr{O}$ with low connected-degree as a class, we mainly discusses θ-classes with high connected-degree. Especially, if $\alpha = 1$ and $\theta = 1$, then a θ-class is essentially a classical class. And further, we can obtain the conclusion that the classical class is only a special case of the θ-class, which also indirectly verifies the reasonability of the classification method based on the relation granularity.

For instance, in Fig. 7, there are following sub-relations

$$1245 \times 1245, 2345 \times 2345, 245 \times 12345,$$
$$368 \times 368, 678 \times 678, 68 \times 3678$$

From above sub-relations we know

$$\{1, 2, 4, 5\}, \{2, 3, 4, 5\}, \{3, 6, 8\}, \{6, 7, 8\}$$

are classical classes. By simple calculation, we can see that $\phi_\alpha(1245 \times 1245) = \phi_\alpha(2345 \times 2345) = \phi_\alpha(368 \times 368) = \phi_\alpha(678 \times 678) = 1$, $\phi_\alpha(245 \times 12345) = 0.6$, $\phi_\alpha(68 \times 3678) = 0.5$, taking $\phi_\alpha(245 \times 12345)$ as an example

$$\phi_\alpha(245 \times 12345) = \frac{|\{2, 4, 5\}|}{|\{2, 4, 5\}| + |\{1, 3\}|} = 0.6$$

when $\alpha = 1, \theta = 0.5$, since $\{3, 6, 7, 8\} = \{6, 8\} \cup \{3, 6, 7, 8\}$ and $\{1, 2, 3, 4, 5\} = \{2, 4, 5\} \cup \{1, 2, 3, 4, 5\}$ are the maximum-subsets satisfying the conditions in Definition 5, they are θ-classes, corresponding results are shown in Fig. 8.

In order to understand θ-class easily, another analytical method is given from the perspective of mutually similar classes.

Definition 6. Let R be a tolerance relation on U, A, A_1, $A_2 \subseteq U$, we define

$$A^+ = \{x \in U \mid (x, y) \in R, \forall y \in A\}$$

if $A_1 \subseteq A_2^+$ and $A_2 \subseteq A_1^+$, then we say A_1 and A_2 are similar to each other; furtherly, if $A_1 = A_2^+$ and $A_2 = A_1^+$, then we say A_2 is a class relative to A_1, and A_1 is a class relative to A_2, in the case, we say A_1 and A_2 are mutually similar classes.

In above definition, if A_1 and A_2 are are mutually similar classes, then we can naturally think about whether the two classes can be merged into one new class.

Theorem 2. *In Definition 6, the following statements are equivalent*

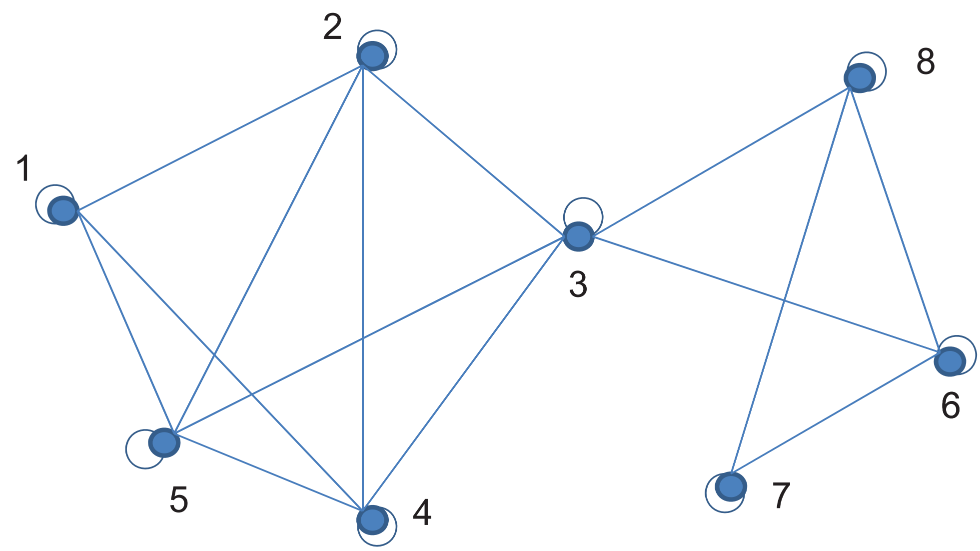

Fig. 7. A tolerance relation graph.

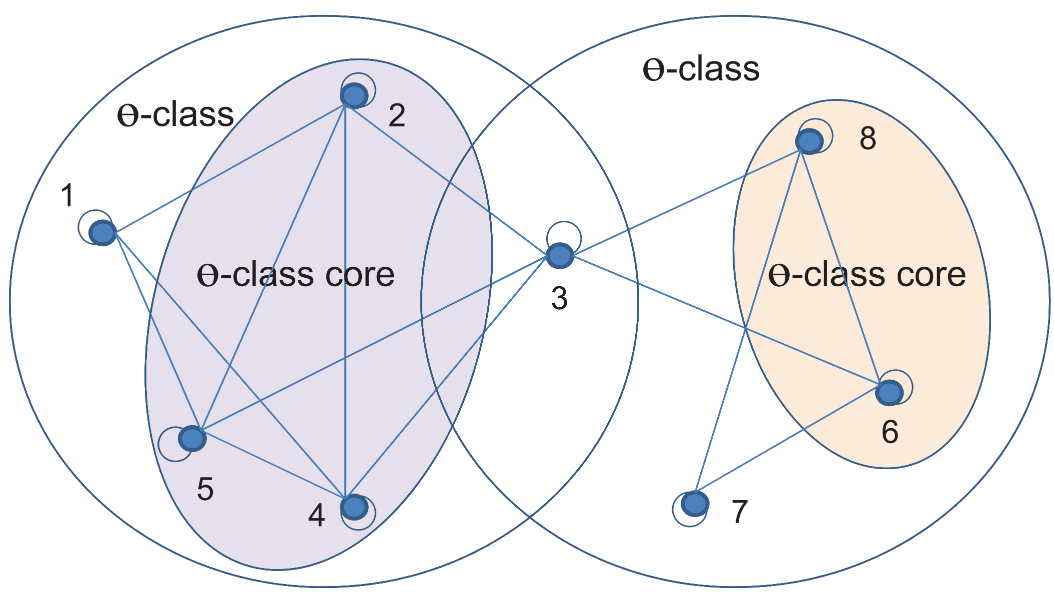

Fig. 8. The θ-classes in Fig. 7 with $\theta = 0.5$ and $\alpha = 1$.

(1) *$A_1 \times A_2$ is a sub-relation in R, namely, $A_1 \times A_2 \in \Sigma(R)$;*
(2) *A_1 and A_2 be mutually similar classes.*

In above theorem, if $\mathscr{O} = A_1 \cup A_2$ is a θ-class, then $\mathscr{O}$ can be viewed as the merging result of mutually similar classes A_1 and A_2 (as shown in Fig. 9).

3.3. One-valued formal contexts derived from information systems

This section mainly presents variable precision relation-concept scales, and finally derives an one-valued formal context from an information system.

Proposition 3. *Let R be a tolerance relation on the set A, then*

$$R = \cup\{X \times X | X \in \Delta(R)\}$$

where $\Delta(R)$ is the set of all classical classes in A.

Definition 7. Let $S_m = (V_m, V_m, I_m)$ be a tolerance scale, the corresponding θ-tolerance scale is defined as

$$S_m^\theta = (V_m, V_m, I_m^\theta)$$

in S_m^θ, I_m^θ is described as

$$I_m^\theta = \cup\{H \times H \,|\, H \in \Delta_\theta(I_m)\}$$

where $\Delta_\theta(I_m)$ is the set of all θ-classes in V_m.

Theorem 3. *From above discussion, it is obvious that $I_m \subseteq I_m^\theta$ is true.*

For example, in Table 2, θ-tolerance scales with respect to numeric attributes b and c are shown in Table 6, where $\alpha = 1$ and $\theta = 0.5$.

Definition 8. Let $S_m^\theta = (V_m, V_m, I_m^\theta)$ be the θ-tolerance scale of $m \in AT$, then an one-valued formal context transformed from $S = (U, AT)$ is defined as

$$K_S^\theta = (U \times U, AT, J_\theta)$$

where J_θ is described as

$$((x, y), m) \in J_\theta \Leftrightarrow (w, v) \in I_m^\theta, \;\; f(x, m) = w, \; f(y, m) = v$$

K_S^θ is called the derivative context of S.

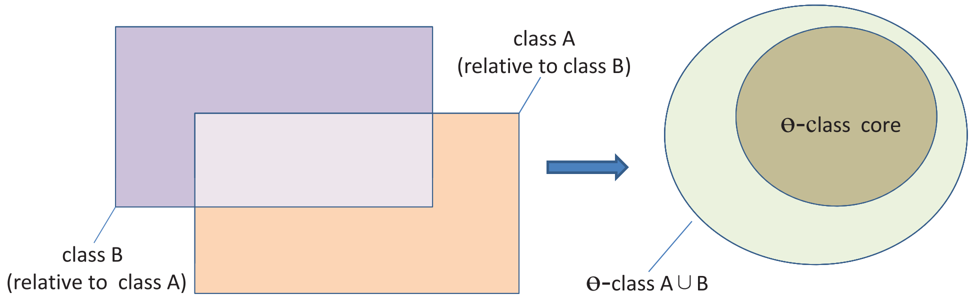

Fig. 9. Generating θ-class from the perspective of mutually similar classes.

Table 6
θ-tolerance scales with respect to attributes b and c in Table 1.

	u_1	u_2	u_3	u_4	u_5	u_6	u_7	u_8		v_1	v_2	v_3	v_4	e_5	v_6	v_7	v_8
u_1	×	×	×	×					v_1	×	×	×	×	×	×		
u_2	×	×	×	×					v_2	×	×	×	×	×	×		
u_3	×	×	×	×	×	×			v_3	×	×	×	×	×	×		
u_4	×	×	×	×	×	×			v_4	×	×	×	×	×	×		
u_5			×	×	×	×			v_5	×	×	×	×	×	×	×	×
u_6			×	×	×	×			v_6	×	×	×	×	×	×	×	×
u_7							×	×	v_7					×	×	×	×
u_8							×	×	v_8					×	×	×	×

Table 7
An one-valued formal context derived from Table 1.

	a	b	c	d
(1, 1)	×	×	×	×
(1, 2)	×	×	×	×
(1, 3)	×	×	×	
(1, 4)	×	×	×	
(1, 5)			×	
(1, 6)			×	
(1, 7)				×
⋮	⋮	⋮	⋮	⋮
(8, 6)	×		×	
(8, 7)	×	×	×	×
(8, 8)	×	×	×	×

By referring to I_a and I_d shown in Table 3, and I_b^θ and I_c^θ shown in Table 6, then an one-valued formal context shown in Table 7 can be derived from Table 1.

4. Variable precision rough set based on θ-concept lattice

Concept lattice, or Galois lattice, as the core data structure in formal concept analysis theory, is a kind of powerful data analysis tool. In this section, an algebraic structure is derived from an information system, which is called θ- concept lattice.

Let R be a tolerance relation on U, then the corresponding cover is denoted as U/R. Let U/R_1 and U/R_2 be covers of U, for any $X \in U/R_1$, if there always exists $Y \in U/R_2$ satisfying $X \subseteq Y$, then we say U/R_2 is coarser than U/R_1 , which is denoted as $U/R_1 \Subset U/R_2$. In the case, we can easily find the following conclusion:

$$R_1 \subseteq R_2, \text{ if and only if } U/R_1 \Subset U/R_2$$

It is known that there always exists a binary relation R_B for any $B \subseteq AT$ in an information system (U, AT). In the paper, by means of S_m^θ, $m \in B$, the corresponding tolerance relation on U relative to B is defined as

$$R_B^\theta = \{(x, y) \in U \times U | \forall m \in B,\ (v, w) \in I_m^\theta,\ f(x, m) = v,\ f(y, m) = w\}$$

In the derivative context K_S^θ, operators can be defined similar to the ones in Definition 2, that is, for any $R \subseteq U \times U$, we define

$$R' = \{m \in AT | ((x, y), m) \in J_\theta, \forall (x, y) \in R\}$$

correspondingly, for any $B \subseteq AT$, we define

$$B' = \{(x, y) \in U \times U | ((x, y), m) \in J_\theta, \forall m \in B\}$$

if $R' = B$ and $B' = R$, then $(U/R, B)$ is called a θ-concept. For any θ-concepts $(U/R_1, B_1)$ and $(U/R_2, B_2)$, we define

$$(U/R_1, B_1) \preccurlyeq (U/R_2, B_2) \Leftrightarrow R_1 \subseteq R_2 \Leftrightarrow U/R_1 \Subset U/R_2 \Leftrightarrow B_2 \subseteq B_1$$

Drawing on the above discussion, it is obvious that $(\mathscr{B}(S_\theta), \preccurlyeq)$ is a complete lattice, which is called θ-concept lattice, where $\mathscr{B}(S_\theta)$ is the set of all θ-concepts. And further, the following conclusions can be immediately.

Theorem 4. *In $K_S^\theta = (U \times U, AT, J_\theta)$, let $B \subseteq AT$, then $B' = R_B^\theta$.*

Theorem 5. *In $K_S^\theta = (U \times U, AT, J_\theta)$, let $B \subseteq AT$, $R \subseteq U \times U$, then*

$$(R, B) \in \mathscr{B}(K_S^\theta) \Leftrightarrow (U/R, B) \in \mathscr{B}(S_\theta)$$

Essentially, as an important algebraic structure in (U, AT), $(\mathscr{B}(S_\theta), \preccurlyeq)$ can organize all hidden covers or partitions of U in the form of lattice structure. There is an indirect way to generate $(\mathscr{B}(S_\theta), \preccurlyeq)$, namely, generating $\mathscr{B}(K_S^\theta)$ by means of classical lattice generating algorithms firstly; secondly, for any concept $(R, B) \in \mathscr{B}(K_S^\theta)$, we replace it by $(U/R, B) \in \mathscr{B}(S_\theta)$, then the replaced lattice is $(\mathscr{B}(S_\theta), \preccurlyeq)$.

In Table 1, when $\alpha = 1$ and $\theta = 1$, then the corresponding $(\mathscr{B}(S_\theta), \preccurlyeq)$ is shown in Fig. 10, when $\alpha = 1$ and $\theta = 0.5$, then the one is shown in Fig. 11. For convenience, any cover or partition $U/R = \{P_1, P_2, \ldots, P_l\}$ in figures is simplified as $P_1/P_2, \ldots, /P_l$, meanwhile, $P_i = \{u_1, u_2, \ldots, u_n\}$ is simplified as $u_1 u_2 \ldots u_n$.

Theorem 6. *In $K_S^\theta = (U \times U, AT, J_\theta)$, let $B, D \subseteq AT$, if $B'' = D$, then*

$$U/R_B = U/R_D$$

where $(U/R_D, D) \in \mathscr{B}(S_\theta)$.

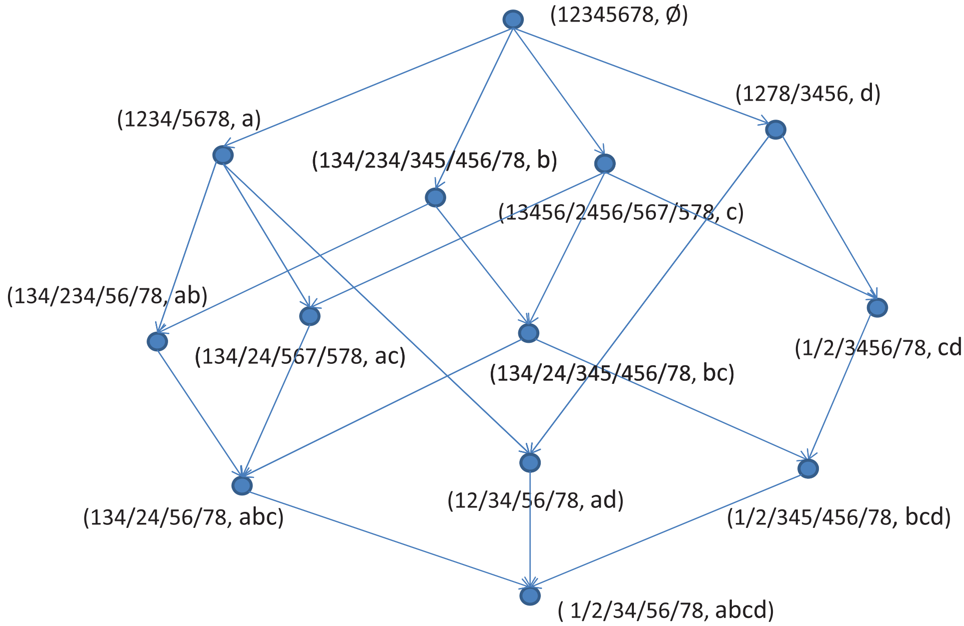

Fig. 10. A θ-concept lattice derived from Table 1 with $\alpha = 1, \theta = 1$.

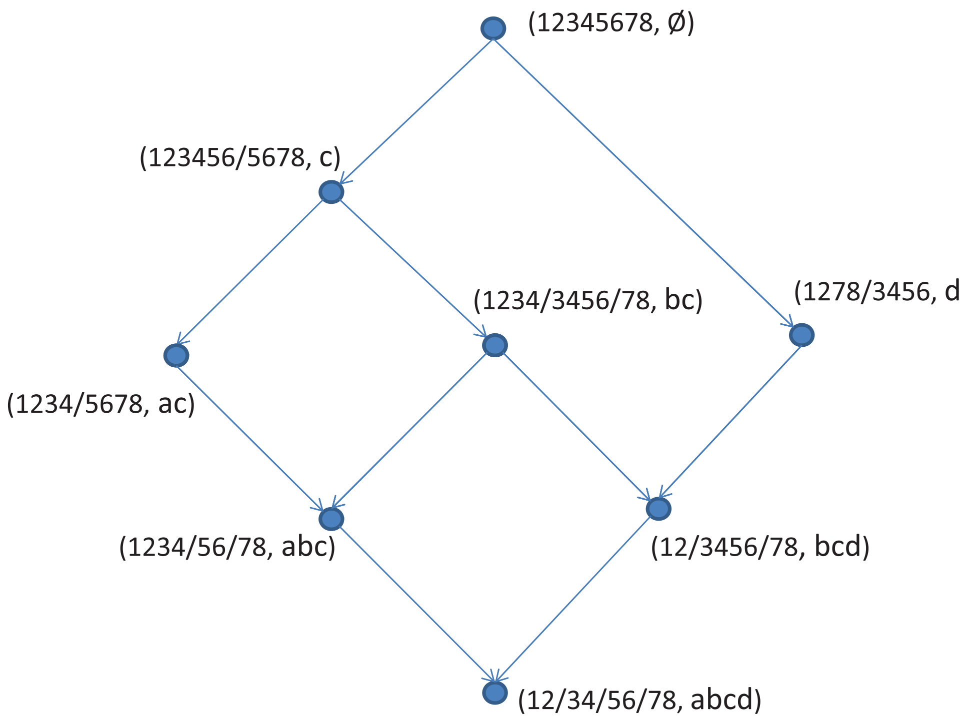

Fig. 11. A θ-concept lattice derived from Table 1 with $\alpha = 1, \theta = 0.5$.

From above theorem, by means of θ-concept, the lower approximation of $X \subseteq U$ relative to $B \subseteq AT$ is defined as

$$\underline{B}(X) = \bigcup_{P \in U/R_D \text{ and } P \subseteq X} P$$

correspondingly, the upper approximation of X relative to B is

$$\overline{B}(X) = \bigcup_{P \in U/R_D \text{ and } P \cap X \neq \varnothing} P$$

where $D = B''$ and $(U/R_D, D) \in \mathscr{B}(S_\theta)$.

Kang et al. have studied the similar problem, namely inducing a lattice structure from an information system [15]. However, this method is always used to process information systems only containing nominal attributes. For instance, based the method in above literature, a concept lattice shown in Fig. 12 can be derived from Table 8, and the corresponding extent of any lattice node is essentially a partition. Comparing with the method in [15], the one presented in the paper not only can obtain the same conclusion from Table 8, but also is valid for information systems containing numeric attributes, the corresponding extent of any lattice node can be a partition or a cover.

Based on above discussions, we can see that the method in the literature [15] is only a special case of the method proposed in the paper essentially.

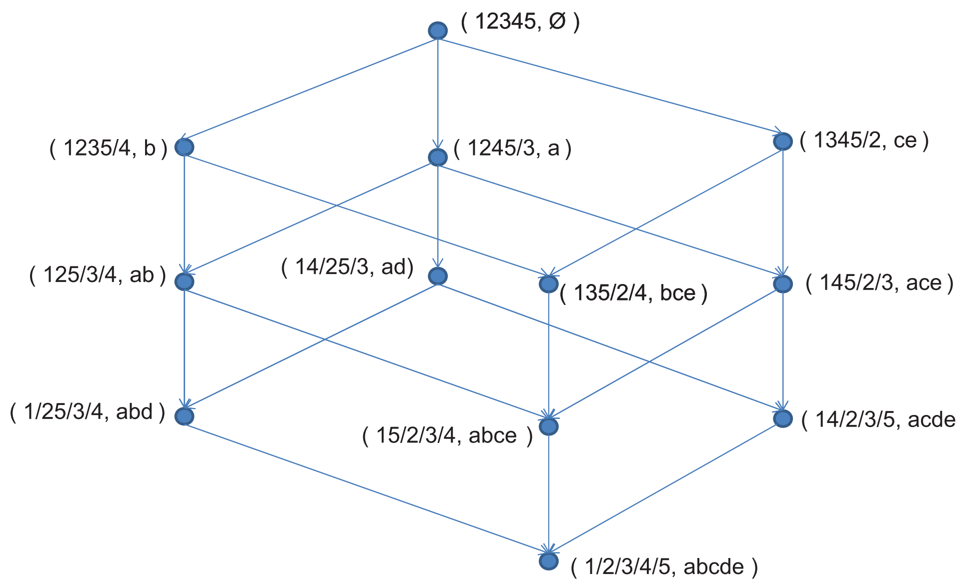

Fig. 12. An algebraic structure derived from Table 8.

Table 8
An information system about cars.

	a	b	c	d	e
1	low	full	diesel	low	low
2	low	full	gasoline	high	high
3	high	full	diesel	medium	low
4	low	compact	diesel	low	low
5	low	full	diesel	high	low

Table 9
Some computing results on the basis of $\mathscr{U}_\theta$ with $\alpha = 1, \theta = 0.5$.

B	$T \in \mathscr{U}_\theta : B \subseteq T$	$\cap\{T \in \mathscr{U}_\theta \mid B \subseteq T\}$
abc	abc, abcd	abc
ab	abc, abcd	abc
ac	ac, abc, abcd	ac
bc	bc, abc, bcd, abcd	bc
a	ac, abc, abcd	ac
b	bc, abc, bcd, abcd	bc
c	c, ac, bc, abc, bcd, abcd	c
Ø	$\mathscr{U}_\theta$	Ø

It is known that information granules, lower and upper approximations are core factors in rough set. In the classical variable precision rough set, the so-called variable precision means approximations are changed by introducing a parameter, rather than information granules are changed. However, in the paper, the variable precision means the information granules are changed by introducing a parameter. Clearly, there exist some differences between two methods, but when people explores approximate solutions rather than accurate ones in some problems, they both contribute enhancing the robustness and generalization ability of rough set.

5. Reduction, core and dependency in information systems

This section offers concept lattice-based solutions to common problems in rough set such as reduction, core and dependency. For above problems, the paper present some relatively simple solutions based on intents of θ-concepts.

In the following theorems, the set of intents of all θ-concepts is denoted as $\mathscr{U}_\theta$.

Theorem 7. *In* $K_S^\theta = (U \times U, AT, J_\theta)$*, let* $m \in B \subseteq AT$*, then* $m \in core(B)$*, if*

$$B \nsubseteq \cap\{T \in \mathscr{U}_\theta \mid (B - m) \subseteq T\}$$

Proof. Since $B \nsubseteq (B-m)''$ can be deduced from Proposition 2 and $B \subseteq B''$, $B'' \nsubseteq (B-m)''$ is true. This implies $B'' \neq (B-m)''$. And further, based on $B'' \neq (B-m)'' \Rightarrow B' \neq (B-m)'$, $R_B^\theta \neq R_{B-m}^\theta$ can be inferred by Theorem 4 easily. Therefore, $m \in core(B)$ holds. □

Theorem 8. *In* $K_S^\theta = (U \times U, AT, J_\theta)$*, let* $C \subseteq B \subseteq AT$*, if C is the minimum-subset satisfying*

$$B \subseteq \cap\{T \in \mathscr{U}_\theta \mid C \subseteq T\}$$

then C is a reduction of B.

Proof. From Proposition 2, it is not hard to see that $B \subseteq C''$ holds. And further, we get $B'' \subseteq C''$. In addition, $C'' \subseteq B''$ can be inferred by $C \subseteq B$ easily. Therefore, $C'' = B''$ is true. Since $C'' = B'' \Rightarrow C' = B' \Rightarrow R_B^\theta = R_C^\theta$ and C is the minimum-subset satisfying $R_B^\theta = R_C^\theta$, C is a reduction of B. □

Theorem 9. *In* $K_S^\theta = (U \times U, AT, J_\theta)$*, let B, C* $\subseteq AT$*, then* $B \to C$*, if*

$$C \subseteq \cap\{T \in \mathscr{U}_\theta \mid B \subseteq T\}$$

Proof. We can see that $C \subseteq B''$ holds by Proposition 2, then $B'' \subseteq C \Rightarrow B' \subseteq C'$ can be deduced. And further, $R_B^\theta \subseteq R_C^\theta$ can be inferred by Theorem 4 easily. Obviously, this implies $B \to C$ holds. □

For example, in Table 1, let $B = \{a, b, c\}$, $\alpha = 1$ and $\theta = 0.5$, by referring Table 9 we can see $core(B) = \{a\}$, $\{a\} \to \{c\}$ is a dependency, and {*a, b*} is a reduction of *B*.

Definition 9. Let (*G, M, I*) be a formal context, *B, D* $\subseteq M$. If $B' \subseteq D'$, then we say $B \to D$ is an implication.

In fact, from the conclusion $R_B^\theta \subseteq R_D^\theta \Leftrightarrow B' \subseteq D'$, we can see that an implication in K_S^θ is essentially a dependency.

Theorem 10. *Let* $B \to D$ *be an implication in* K_S^θ*, if* $B \subseteq B_1$ *and* $D_1 \subseteq D$*, then* $B_1 \to D_1$*.*

Proof. It is obvious that $B' \subseteq D'$ can be inferred from $B \to D$. In addition, $B'_1 \subseteq B'$ and $D' \subseteq D'_1$ can be deduced by $B \subseteq B_1$ and $D_1 \subseteq D$ separately. Hence $B'_1 \subseteq D'_1$ is true, that is, $B_1 \to D_1$ holds. □

Table 10
The set of all core dependencies in Table 1 with $\alpha = 1$ and $\theta = 1$.

$bd \rightarrow bcd$	$ab \rightarrow ab$	$cd \rightarrow cd$	$d \rightarrow d$
$b \rightarrow b$	$abd \rightarrow abcd$	$abc \rightarrow abc$	$\emptyset \rightarrow \emptyset$
$acd \rightarrow abcd$	$ad \rightarrow ad$	$a \rightarrow a$	$ac \rightarrow ac$
$c \rightarrow c$	$bc \rightarrow bc$		

Table 11
The set of all core dependencies in Table 1 with $\alpha = 1$ and $\theta = 0.5$.

$ab \rightarrow abc$	$b \rightarrow bc$	$a \rightarrow ac$
$d \rightarrow d$	$ad \rightarrow abcd$	$c \rightarrow c$
$bd \rightarrow bcd$	$cd \rightarrow bcd$	$\emptyset \rightarrow \emptyset$

We know that dependency, as a kind of special rule-type knowledge, is a common knowledge expression form for its advantages such as strong description and easy understanding. Concept lattice can organize data in the form of lattice, can manifest the "generalization and specialization" relations among concepts, so it is suitable to discover rules. In fact, $(\mathscr{B}(S_\theta), \preccurlyeq)$, as a kind of special concept lattice, is suitable for mining rule-type knowledge as well. Although concept lattice can be applied to mining dependencies, the number of dependencies acquired is big, and there are a mass of redundant dependencies. In the case, we can further eliminate some common redundant dependencies by means of some inference-rule, and finally obtain a smaller set of dependencies. At present, there have been many findings on inference-rule. In the following paper, we chose a relatively easy inference-rule, which is deduced from Theorem 10.

Inference-rule: If $B \subseteq B_1$ and $D_1 \subseteq D$, then $B_1 \rightarrow D_1$ can be inferred from $B \rightarrow D$. In fact, the inference-rule can also be represented as

$$\frac{B \subseteq B_1,\ D_1 \subseteq D,\ B \rightarrow D}{B_1 \rightarrow D_1}$$

Inspired by previous research results [15,21], we define core dependencies as follows:

Definition 10. In $K_S^\theta = (U \times U, AT, J_\theta)$, let $B \subseteq AT$, if D is a reduction of some subset in AT, then we say $D \rightarrow D''$ is a core dependency.

Theorem 11. *For any $B \rightarrow C$, if D is a reduction of B, then $B \rightarrow C$ can be inferred from $D \rightarrow D''$ by means of the inference-rule.*

Proof. Since D is a reduction of B, $D \subseteq B$ and $R_B^\theta = R_D^\theta \Rightarrow B' = D'$ can be obtained easily. In addition, $B' \subseteq C'$ can be deduced from $B \rightarrow C$. Therefore $D' \subseteq C'$ holds, and then $C'' \subseteq D''$ can be obtained. Together with $C \subseteq C''$ we can get $C \subseteq D''$. In the case, we can see that $D \rightarrow D''$ satisfies $D \subseteq B$ and $C \subseteq D''$. Obviously, $B \rightarrow C$ can be inferred from $D \rightarrow D''$ by the inference-rule. □

In above theorem, since D is a reduction of B, $D \rightarrow D''$ is a core dependency. It is not hard to see that any $B \rightarrow C$ can be inferred from core dependencies by the inference-rule.

Based on the points discussed above, we can provide users with a small set of dependencies, and users can selectively derive other dependencies from the core dependencies according to their interests. For instance, in Table 1, when $\alpha = 1, \theta = 1$, then the set of all core dependencies is shown in Table 10; when $\alpha = 1, \theta = 0.5$, then the set of all core dependencies is shown in Table 11. When $\alpha = 1, \theta = 0.5$, taking $abd \rightarrow c$ as an example, we know $\{a, d\}$ is a reduction of $\{a, b, d\}$, and $\{a, d\}'' = \{a, b, c, d\}$, since $\{a, d\} \subseteq \{a, b, d\}$ and $\{c\} \subseteq \{a, b, c, d\}$, $\{a, b, d\} \rightarrow \{c\}$ can be inferred from $\{a, d\} \rightarrow \{a, b, c, d\}$ by the inference-rule.

6. Summary and outlook

Concept lattice is a kind of mathematical tool for data analysis and processing. As two mathematical branches generated in the same era, concept lattice and rough set vary from each other in their research methods, but the same research background and objective indicate that they must have something in common. Constructing the connection between two theories via their respective advantages is conducive to abstracting a more general and universal data analysis framework. In the case, the paper brings concept lattice into rough set with a view to its advantages such as outstanding mathematical property, intuitive lattice structure, abundant semantics of concepts, and so on.

To further enhance the robustness and generalization ability of rough set based on the tolerance relation, this paper, by mean of the granularity of tolerance relation, expands classical classes to θ-classes, which actually just requires "most" rather than "all" elements in one class are similar to each other. The so-called granularity of tolerance relation means that the tolerance relation can be decomposed into several strongly connected sub-relations and several weakly connected sub-relations. And further, the paper emphatically probes the θ-concept lattice $(\mathscr{B}(S_\theta), \preccurlyeq)$ inferred from an information system. As an important algebraic structure, $(\mathscr{B}(S_\theta), \preccurlyeq)$ can organically organize all covers or partitions of U in the form of lattice structure. Meanwhile, the paper also offers solutions based concept lattice to common problems in rough set, such as attribute reduction, core and dependency.

Both theories and examples demonstrate that the conclusion of the paper is reasonable and valid. Obviously, introducing concept lattice into the study of rough set is an effective means to expand rough set. In short, the paper not only can be viewed as an useful exploration and attempt for the fusion study of the two theories, but also offers a new idea for the expansion of rough set. Furthermore, it also helps to reasonably analyze and interpret rough set from the perspective of concept lattice. Although the paper has proposed some significant theoretical findings, these findings have to be further supplemented and improved. Our research focuses in the next step will include how to further apply concept lattice to deal with more complicated information systems, how to effectively reduce the time complexity and space complexity in the process of knowledge acquisition based on concept lattice, and how to further improve the fusion theory of two theories and finally abstract a more common and universal data analysis tool, etc.

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] R. Belohlavek, A note on variable threshold concept lattices: threshold-based operators are reducible to classical concept forming operators, Inf. Sci. 177 (2007) 3186–3191.
[2] 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.
[3] J.H. Dai, Rough set approach to incomplete numerical data, Inf. Sci. 241 (2013) 43–57.
[4] J.S. Deogun, J. Saquer, Monotone concepts for formal concept analysis, Discrete Appl. Math. 144 (2004) 70–78.
[5] I. Duntsch, G. Gediga, Approximation operators in qualitative data analysis, Theory and Application of Relational Structures as Knowledge Instruments, Springer, Heidelberg, 2003, pp. 216–233.
[6] S.Q. Fan, W.X. Zhang, W. Xu, Fuzzy inference based on fuzzy concept lattice, Fuzzy Set. Syst. 157 (2006) 3177–3187.

[7] B. Ganter, R. Wille, Formal Concept Analysis: Mathematical Foundations, Springer-Verlag, Berlin, 1999.
[8] G. Gediga, I. Duntsch, Modal-style operators in qualitative data analysis, in: Proceedings of the 2002 IEEE International Conference on Data Ming, 2002, pp. 155–162.
[9] S. Greco, B. Matarazzo, R. Slowinski, Rough approximation of a preference relation by dominance relation, Euro. J. Oper. Res. 117 (1999) 63–83.
[10] Y.Y. Guan, H.K. Wang, Set-valued information systems, Inf. Sci.s 176 (17) (2006) 2507–2525.
[11] J. Hu, T.R. Li, H.J. Wang, H. Fujita, Hierarchical cluster ensemble model based on knowledge granulation, Knowl. Based Syst. 91 (2016) 179–188.
[12] Q.H. Hu, D.R. Yu, J.F. Liu, C.X. Wu, Neighborhood rough set based heterogeneous feature subset selection, Inf. Sci. 178 (2008a) 3577–3594.
[13] Q.H. Hu, J.F. Liu, D.R. Yu, Mixed feature selection based on granulation and approximation, Knowl. Based Syst. 21 (4) (2008b) 294–304.
[14] X.P. Kang, D.Y. Li, S.G. Wang, K.S. Qu, Formal concept analysis based on fuzzy granularity base for different granulations, Fuzzy Set. Syst. 203 (2012) 33–48.
[15] 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.
[16] 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.
[17] R.E. Kent, Rough concept analysis, Fund. Inform. 27 (1996) 169–181.
[18] 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.
[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.H. Li, C.L. Mei, Y.J. Lv, Knowledge reduct in decision formal contexts, Knowl. Based Syst. 24 (5) (2011) 709–715.
[21] 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.
[22] 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.
[23] J.H. Li, C.L. Mei, Y.J. Lv, Incomplete decision contexts: approximate concept construction, rule acquisition and knowledge reduction, Int. J. Approx. Reason. 54 (2013) 149–165.
[24] T.J. Li, W.Z. Wu, Attribute reduction in formal contexts: a covering rough set approach, Fund. Inform. 111 (1) (2011) 15–32.
[25] 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.
[26] 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.
[27] 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.
[28] J.M. Ma, Y. Leung, W.X. Zhang, Attribute reductions in object-oriented concept lattices, Int. J. Mach. Learn. Cybern. 5 (2014) 789–813.
[29] 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.
[30] G.D. Oosthuizen, Rough sets and concept lattices, in: W.P. Ziarko (Ed.), Rough Sets, and Fuzzy Sets and Knowledge Discovery (RSKD'93), Springer-Verlag, London, 1994, pp. 24–31.
[31] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (1982) 341–356.
[32] J.J. Qi, T. Qian, L. Wei, The connections between three-way and classical concept lattices, Knowl. Based Syst. 91 (2016) 143–151.
[33] Y.H. Qian, J.Y. Liang, W. Pedrycz, C.Y. Dang, An efficient accelerator for attribute reduction from incomplete data in rough set framework, Pattern Recogn. 44 (8) (2011) 1658–1670.
[34] Y.H. Qian, J.Y. Liang, D.Y. Li, F. Wang, N.N. Ma, Approximation reduction in inconsistent incomplete decision tables, Knowl. Based Syst. 23 (5) (2010) 427–433.
[35] J.Y. Shyng, H.M. Shieh, G.H. Tzeng, An integration method combining rough set theory with formal concept analysis for personal investment portfolios, Knowl. Based Syst. 23 (6) (2010) 586–597.
[36] J. Saquer, J.S. Deogun, Concept approximations based on rough sets and similarity measures, Int. J. Appl. Math. Comput. Sci. 11 (3) (2001) 655–674.
[37] A.K. Sarmah, S.M. Hazarika, S.K. Sinha, Formal concept analysis: current trends and directions, Artif. Intell. Rev. 44 (1) (2015) 47–86.
[38] M.W. Shao, Y. Leung, Relations between granular reduct and dominance reduct in formal contexts, Knowl. Based Syst. 65 (2014) 1–11.
[39] M.W. Shao, M. Liu, W.X. Zhang, Set approximations in fuzzy formal concept analysis, Fuzzy Set. Syst. 158 (23) (2007) 2627–2640.
[40] A. Skowron, J. Stepaniuk, Tolerance approximation spaces, Fund. Inform. 27 (2) (1996) 245–253.
[41] R. Slowinski, D. Vanderpooten, Similarity relation as a basis for rough approximations, in: P. Wang (Ed.), Advances in Machine Intelligence and Soft-Computing, 4, University Press, Duke, 1997, pp. 17–33.
[42] R. Slowinski, D. Vanderpooten, A generalized definition of rough approximations based on similarity, IEEE Trans. Data Knowl. Eng. 12 (2) (2000) 331–336.
[43] 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.
[44] L.D. Wang, X.D. Liu, Concept analysis via rough set and AFS algebra, Inf. Sc. 178 (21) (2008) 4125–4137.
[45] L. Wei, J.J. Qi, Relation between concept lattice reduct and rough set reduct, Knowl. Based Syst. 23 (8) (2010) 934–938.
[46] R. Wille, Restructuring lattice theory: an approach based on hierarchies of concepts, in: I. Rival (Ed.), Ordered Sets, Reidel, Dordrecht, 1982, pp. 445–470.
[47] Q. Wu, Z.T. Liu, Real formal concept analysis based on grey-rough set theory, Knowl. Based Sys. 22 (1) (2009) 38–45.
[48] W.H. Xu, W.T. Li, Granular computing approach to two-way learning based on formal concept analysis in fuzzy datasets, IEEE Trans. Cybern. 46 (2) (2016) 366–379.
[49] Y.Y. Yao, comparative study of formal concept analysis and rough set theory in data analysis, in: Proceedings of the 3rd International Conference on Rough Sets and Current Trends in Computing, RSCTC'04, LNCS, 2004, pp. 59–68.(3066).
[50] Y.Y. Yao, On unifying formal concept analysis and rough set analysis, in: W.X. Zhang, Y.Y. Yao, Y. Leung (Eds.), Rough Set and Concept Lattice, Xi'an Jiaotong University Press, Xi'an, 2006, pp. 1–20.
[51] Y.Y. Yao, Y.H. Chen, Rough set approximations in formal concept analysis, LNCS Trans. Rough Sets V 4100 (2006) 285–305.
[52] Y.H. Zhai, D.Y. Li, K.S. Qu, Decision implication canonical basis: a logical perspective, J. Comput. Syst. Sci. 81 (2015) 208–218.
[53] W.X. Zhang, J.M. Ma, S.Q. Fan, Variable threshold concept lattices, Inf. Sci. 177 (2007) 4883–4892.