Front. Electr. Electron. Eng. China 2008, 3(4): 399–405
DOI 10.1007/s11460-008-0070-y

RESEARCH ARTICLE

Guoyin WANG, Lihe GUAN, Feng HU

Rough set extensions in incomplete information systems

Abstract All eight possible extended rough set models in incomplete information systems are proposed. By analyzing existing extended models and technical methods of rough set theory, the strategy of model extension is found to be suitable for processing incomplete information systems instead of filling possible values for missing attributes. After analyzing the definitions of existing extended models, a new general extended model is proposed. The new model is a generalization of indiscernibility relations, tolerance relations and non-symmetric similarity relations. Finally, suggestions for further study of rough set theory in incomplete information systems are put forward.

Keywords rough set, incomplete information system, extended model

1 Introduction

In considering the existence of missing attribute values in an information system (IS), ISs can be divided into complete information systems (CIS) and incomplete information systems (IIS). There has been significant progress in exploring CISs such as rough set theory, evidence theory, and fuzzy sets. The rough set theory developed by Pawlak in the 1980s [1] is an efficient approach of processing imprecision, vagueness and uncertainty often used in data mining, machine learning, and knowledge discovery [2].

Translated from *Progress of Artificial Intelligence in China*, 2007, 980–985 [译自: 中国人工智能进展]

Guoyin WANG (✉), Lihe GUAN, Feng HU
School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, China
E-mail: wanggy_cupt@yahoo.com.cn

Guoyin WANG, Feng HU
Institute of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Lihe GUAN
Institute of Information and Calculation Science, Chongqing Jiaotong University, Chongqing 400074, China

The classical rough set theory is based on the assumption that all objects have deterministic values on every attribute and classifications are made by an indiscernibility relation (or equivalence relation). However, it is not always possible to define an indiscernibility relation in IISs because of missing attribute values. The classical rough set theory can process CISs only, but not IISs directly. Its application is limited to a significant degree. Since some data could not be obtained for various reasons (e.g., capacity, technology, financing), many information systems are always incomplete. Therefore, it is important to study the rough set theory in IISs. It is a challenge for researchers to design sophisticated learning algorithms to process IISs.

It is more difficult to generate knowledge from IISs than from CISs. This problem has attracted many researchers' attention in recent years. For example, Stefanowski [3], Wang [4], Grzymala-Busse [5], et al. analyzed the semantics of missing attribute values in IISs. Kryszkiewicz studied the extraction of association rules from IISs without decision attributes [6]. There are usually two strategies in rough set theory to process IISs: data reparation [4,7] and model extension [3,5,8–14]. The strategy of data reparation is an indirect method that transforms an IIS into a CIS according to rules (usually probability statistical methods), where we can acquire knowledge with the classical rough set theory. However, this strategy changes the original information of IISs and the knowledge systems generated lacks objectiveness. The second strategy, model extension, is a direct method that extends basic concepts of the classical rough set theory in IISs by relaxing the requirement of indiscernibility relation of reflexivity, symmetry and transitivity, i.e., the indiscernibility relation is extended to inequivalence relations that can process IISs directly. For example, Kryszkiewicz [8] proposed a tolerance relation based on Grzymala-Busse's [12] work; Stefanowski and Tsoukiàs developed a non-symmetric similarity relation [3,9]; Wang proposed a limited tolerance relation [11]; Grzymala-Busse proposed a characteristic relation [5,13]. In this paper, an analysis of the two technical strategies shows that the strategy of model extension is suitable for processing IISs. All eight possible extended models of the rough set theory are proposed. Four possible extended models that dissatisfy reflexive properties are further discussed. By analyzing the definitions of existing extended models, a new general extended model, that generalizes the indiscernibility relation, tolerance relation and non-symmetric similarity relation is proposed. Finally, problems for further study of the rough set theory in IISs are proposed.

2 Methods for processing IISs

2.1 Data reparation

There are two major methods for data reparation: delete objects with missing attribute values and transform an IIS into a CIS. Although the first method is not a real method of data reparation, we still refer to it as data reparation. Some main algorithms for data reparation are shown in Table 1.

Table 1 Methods of data reparation

No.	methods
1	most common attribute value
2	concept of most common attribute value
3	assigning all possible values of the attribute
4	assigning all possible values of the attribute restricted to the given concept
5	ignoring examples with missing attribute values
6	treating missing attribute values as special values
7	event-covering method
8	a special LEM2 algorithm
9	C4.5

In Ref. [7], Grzymala-Busse compared nine different algorithms for data reparation. Ten input data files were used to investigate the performance of the nine algorithms to missing attribute values. Setting the average error rate as the quality criterion, both the C4.5 approach and the method of ignoring examples with missing attribute values are the best, while the method of the most common attribute value is the worst. We think that the quality of each data reparation algorithm is related to the IISs. Therefore, these results reported by Grzymala-Busse in limited experimental data do not have enough evidence to support the claim that these approaches are superior.

The algorithms for data reparation in Table 1 can be divided into four types: deleting approach, exhaustion approach, special value approach and statistical approach. The deleting approach loses the original information. The exhaustion approach expands the original data and results in an NP problem. The special value approach takes missing attribute values as special values, which are quite different from any other known attribute values and lacks rationality. The statistical approach adopts various principles to estimate the missing attribute values and inevitably leads to new contradictions and changes the original information. We can conclude that some approaches of data reparation are very convenient for application. However, regardless of the approach used, an artificial estimation of the missing attribute values in IISs is still derived, and the original information of IISs will be changed. Hence, the results generated using these approaches cannot exactly reflect the original information of IISs.

2.2 Model extension

Much work has been done to keep the information of IISs unchanged in data mining by extending the concepts of the classical rough set theory. Several extended models are proposed (e.g., tolerance relation, non symmetric similarity relation, valued tolerance relation, limited tolerance relation, and characteristic relation), which can be used to process IISs directly, and their results are more objective.

2.2.1 Eight possible extended models

The indiscernibility relation is the most basic concept of the classical rough set theory. It is reflexive, symmetric, transitive, and is also an equivalence relation. In IISs, the major issue for extending the classical rough set theory is to extend the indiscernibility relation according to inequivalence relations. According to the requirement of reflexivity, symmetry and transitivity, all possible extended models can be classified into eight types, as shown in Fig. 1. All the existing typical extensions of the rough set theory listed in Table 2 are reflexive.

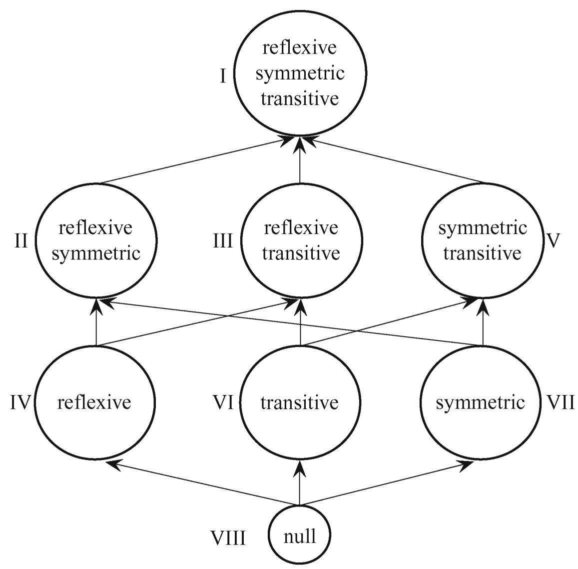

Fig. 1 Eight kinds of possible extended models of rough set theory

Table 2 Category of existing extended models

category	reflexive	symmetric	transitive	existing extended models
I	✓	✓	✓	indiscernibility relation (I)
II	✓	✓	✗	tolerance relation (T), limited tolerance relation (L)
III	✓	✗	✓	non-symmetric similarity relation (S)
IV	✓	✗	✗	characteristic relation
V	✗	✓	✓	
VI	✗	✗	✓	
VII	✗	✓	✗	
VIII	✗	✗	✗	

Note: "✗" is dissatisfied, "✓" is satisfied

2.2.2 Four Categories of existing extended rough set models

The five existing extended models belong to the first four categories as follows:

1) The 1st category models

The 1st category of extended models is the strictest and require reflexive, symmetric and transitive properties. This category model can only be used to process CISs instead of IISs directly. The indiscernibility relation of the classical rough set theory proposed by Pawlak belongs to this category.

2) The 2nd category models

The 2nd category of extended models needs to be reflexive and symmetric, but not necessarily transitive. The tolerance relation and limited tolerance relation belong to this category. In the tolerance relation proposed by Kryszkiewicz, there is an assumption that missing attribute values may be (i.e., equal to) any known attribute values in IISs. Such an interpretation corresponds to the idea that all missing attribute values are just "do not care" conditions [5]. Although this assumption ensures mining as much knowledge as possible, it has the following shortcomings:

i) It makes the largest range of missing attribute values.

ii) It makes the algorithm more difficult and complicated.

iii) It cannot process IISs where all missing attribute values are "lost" values.

iv) The tolerance degree of the objects described by the tolerance relation is 0 or 1.

The tolerance relation is so relaxed that some objects discerned intuitively cannot be classified. Therefore, Stefanowski extended the range of the objects' tolerance degree to [0,1], and proposed a valued tolerance relation, where the probability distribution of information system should be known in advance. Unfortunately, it is very difficult for a new IIS to achieve. Since the total information of IISs is hard to establish, determining its accurate probability distribution is even more difficult. Hence, its application is limited [4].

The limited tolerance relation inherits the merits of tolerance relation and non-symmetric similarity relation and discards their limitations. However, it does not allow for any error. Although some objects' values are equal in a majority of attributes and not equal only for one or two, they are grouped into different tolerance classes. In fact, it is accepted that there are some errors considered as noise data in exceptional situations.

3) The 3rd category models

The 3rd category of extended models needs to be reflexive and transitive, but not necessarily symmetric. The non-symmetric similarity relation proposed by Stefanowski and Tsoukiàs belongs to this category. Compared with the tolerance relation, the non-symmetric similarity relation takes all missing attribute values in IISs as "lost" values. It cannot process IISs where all missing attribute values are "do not care" conditions.

4) The 4th category models

The 4th category of extended models needs to be reflexive, but not necessarily symmetric or transitive. The characteristic relation proposed by Grzymala-Busse belongs to this category. Compared with the tolerance relation and the non-symmetric similarity relation, the characteristic relation can process IISs with both types: lost values and "do not care" conditions [5]. However, this relation is only a simple syncretism for the tolerance relation and the non-symmetric similarity relation and cannot avoid their shortcomings.

In Ref. [11], an attribute subset B is given in IISs. The inclusion relations of lower and upper approximation sets of an object subset X defined by the tolerance relation, the non-symmetric similarity relation and the limited tolerance relation are discussed. From the approximate measure of X, the non-symmetric similarity relation is the strictest, the tolerance relation is the most relaxed, and the limited tolerance relation is in between.

2.3 Comparison and analysis of two strategies

The data reparation is a strategy of assigning an appropriate known attribute value to the missing attribute value in particular ways or ignoring objects with missing attribute values, and transforming an IIS into a CIS. This process is also called the completeness of IISs. Then, reduce and acquire knowledge in CISs with the classical rough set theory. The model extension is another strategy of extending the indiscernibility relation to the inequivalence relation, and any information from the original IIS will be unchanged in the knowledge generation process. The basic processes of these two different strategies of knowledge acquisition are shown in Fig. 2.

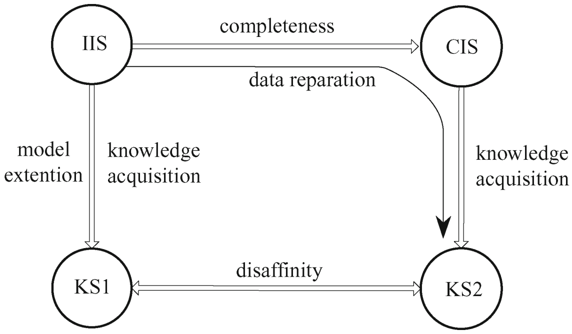

Fig. 2 Comparison of two strategies about knowledge acquisition in IISs

From the process of knowledge acquisition shown in Fig. 2, data reparation has two steps (completeness and knowledge acquisition) to obtain the knowledge system KS2. In Sect. 2.1, we concluded that the main shortcoming of this strategy is that it changes the information of the original information system, i.e., CIS $\neq$ IIS. However, the model extension processes IISs directly, which avoids the shortcomings of data reparation. There are some differences between the knowledge in IISs and that in CISs because the original information of IISs is changed in the completeness process. The knowledge systems KS1 and KS2 are different, and KS1 is more impersonal in maintaining the knowledge in IISs than KS2. Therefore, based on the idea that the process of knowledge acquisition is the process of transforming knowledge format [14], the model extension is superior to data reparation.

3 Analysis of four possible non-reflexive extended models

The existing tolerance relation, non-symmetric similarity relation, the limited tolerance relation, and the characteristic relation are all reflexive and belong to the first four categories of extended models, while the last four categories of extended models in Table 2 are used infrequently. However, many relations are not reflexive in reality. For example, the relation between subordinates and supervisors is transitive, but not reflexive or symmetric, and it belongs to the 6th category. The opposite relation is symmetric, but not reflexive or transitive and belongs to the 7th category. However, is it possible to define a non-reflexive relation in IISs? Is it helpful to knowledge acquisition? What is its characteristic? In the following sections, for the convenience of discussion and analysis, the set of all objects is denoted by U, and the set of all condition attributes is denoted by $C=\{c_1,c_2,c_3,c_4\}$. Let b be an attribute, i.e., $b\in C$ and the domain of b be denoted by V_b. The non-symmetric similarity relation S_b is defined as

$$S_b=\{(x,y)\in U\times U|b(x)=*\vee b(x)=b(y)\}.$$

In IISs, there are five cases for the attribute values of any two objects x and y in U on b:

1) $b(x)=*,\ b(y)=*$; 2) $b(x)=*,\ b(y)\neq *$;

3) $b(x)\neq *,\ b(y)=*$; 4) $b(x)=b(y)\neq *$;

5) $b(x)\neq *,\ b(y)\neq *,\ \ b(x)\neq b(y)$.

If the probability distribution of all attribute values in V_b is a uniform distribution, the similarity probability of x and y is always $1/|V_b|$ with reference to b in the first three cases. They satisfy the tolerance relation on b. In the first two cases, $(x,y)\in S_b$. In the 3rd case, $(x,y)\notin S_b$ and $(y,x)\in S_b$. Here, when the similarity probability of two objects is equal on b, they satisfy the tolerance relation, but do not always satisfy the non-symmetric similarity relation. Thus, the non-symmetric similarity relation is stricter than the tolerance relation.

On the other hand, all missing attribute values may be taken as any known attribute values in the tolerance relation, and the non-symmetric similarity relation assumes that missing attribute values cannot be compared. For example, $x=\{1,2,3,*\}$ and $y=\{1,2,3,*\}$ satisfy the tolerance relation and the non-symmetric similarity relation on C. Evidently, this is not reasonable (we cannot exclude the difference of their values on c_4, i.e., they may be discerned only by c_4. In this case, they cannot be put into the same class). The tolerance relation and the non-symmetric similarity relation are two extremities in this case. Stefanowski and Tsoukiàs proposed a valued tolerance relation by measuring the similarity degree of the objects [10]. This model defines the tolerance class by pre-establishing a parameter λ, which overcomes the shortcomings of the tolerance relation and the non-symmetric similarity relation. For a special IIS, the valued tolerance relation is reflexive, symmetric and transitive depending on the parameter λ. For example, if the similarity degree of x and y is 0.6, they satisfy the valued tolerance relation when λ is 0.5, but dissatisfy the valued tolerance relation when λ is 0.7. Hence, the valued tolerance relation is not reflexive at this time.

Based on the above discussion, it could be established that when the attribute values of x and y are all missing values on some attributes, the processing approaches of the tolerance relation and the non-symmetric similarity relation are extreme. The valued tolerance relation is not always reflexive. When the attribute value of x or y is missing on b, and all the values in V_b are a uniform distribution, the probabilistic degree that x and y cannot be discerned on b is $1/|V_b|$. In this case, the greater $|V_b|$ is, the lower the similarity degree of x and y will be on b. We may figure out the definitions of some possible non-reflexive extended models missing in Table 2.

The first relation is defined as follows:

$$R_B^1 = \{(x,y) | x,y \in U, \forall_{b \in B}(b(x) = b(y) \neq *)\},$$

where $B \subseteq C$. Obviously, R_B^1 is symmetric and transitive but not reflexive. Therefore, it belongs to the 5th category of extended models. In fact, when all attribute values of two objects are all known values and identical to each other, they satisfy R_B^1.

The second relation is defined as follows:

$$R_B^2 = \{(x,y) | x,y \in U, (P_B(x) \cap P_B(y) = \varnothing) \wedge (\forall_{b \in B}(b(x) \neq * \rightarrow b(x) = b(y)))\},$$

where $B \subseteq C$, $P_B(x) = \{b | b \in B \wedge (b(x) = *)\}$. Obviously, R_B^2 is transitive, but not reflexive or symmetric. It belongs to the 6th category of the extended models. In fact, an object with missing attribute values is only possible similar to the objects without any missing attribute values.

The third relation is defined as follows:

$$R_B^3 = \{(x,y) | x,y \in U, (P_B(x) \cap P_B(y) = \varnothing) \wedge (\forall_{b \in Q_B(x) \cap Q_B(y)}(b(x) = b(y))\},$$

where $B \subseteq C$, $P_B(x) = \{b | b \in B \wedge (b(x) = *)\}$, $Q_B(x) = B - P_B(x)$. R_B^3 is clearly symmetric, but not reflexive or transitive. It belongs to the 7th category of extended models. According to the definition of R_B^3, when two objects' attribute values are not missing at the same time on any attribute, and their values are equal on all attributes in which their values are all known, they satisfy R_B^3.

The 8th category of the extended model does not need to be reflexive, symmetric and transitive. This category model is an ordinary binary relation.

The 3 kinds of relations proposed above are not reflexive, which fill the extended models missing in Table 2.

4 Novel general extended model

In rough set theory, knowledge is the ability to classify. For any two objects $x,y \in U$, they may be discerned by their attribute values, i.e., their similarity degree should be considered. The higher the similarity degree is, the more difficult for the two objects to be discerned. Thus, Stefanowski and Tsoukiàs proposed a valued tolerance relation [10]. However, there is only one parameter $\lambda \in [0,1]$ in their model, and this extended model is very different from the other models. In this section, a new valued tolerance relation with six parameters is proposed. In the proposed relation, different values of the parameters form different extended models, i.e., all the existing extended models could be taken as special cases.

In IISs, given any object $x,y \in U$ and any attribute $b \in C$, the five cases for the attribute values of x and y are shown in Sect. 3. Therefore, the similarity degree $d(x,y)$ of x and y could be defined as follows:

$$d(x,y) = \frac{|K_1| \cdot \lambda_1 + |K_2| \cdot \lambda_2 + |K_3| \cdot \lambda_3 + |K_4| \cdot \lambda_4 + |K_5| \cdot \lambda_5}{|C|},$$

where $K_1 = \{b \in C | b(x) = *, b(y) = *\}$, $K_2 = \{b \in C | b(x) = *, b(y) \neq *\}$, $K_3 = \{b \in C | b(x) \neq *, b(y) = *\}$, $K_4 = \{b \in C | b(x) = b(y) \neq *\}$, $K_5 = \{b \in C | b(x) \neq *, b(y) \neq *, b(x) \neq b(y)\}$, $C = \bigcup_{i=1}^{5} K_i$, $K_i \cap K_j = \varnothing$, $i \neq j$, $i,j = 1,2,3,4,5$. The five parameters $\lambda_i \in [0,1]$, $i = 1,2,3,4,5$, need to be set up in advance. Evidently, $d(x,y) \in [0,1]$. The new valued tolerance relation R is defined as follows:

$$R = \{(x,y) | x,y \in U, d(x,y) \geqslant \lambda\},$$

where the parameter $\lambda \in [0,1]$ also needs to be set up in advance. For any object $x \in U$, the tolerance class $[x]_R$ is defined as follows:

$$[x]_R = \{y \in U | (x,y) \in R\} = \{y \in U | d(x,y) \geqslant \lambda\}.$$

Let X be a subset of the universe U of all objects. The upper approximation $\bar{X}_B^R$ and the lower approximation $\underline{X}_B^R$ of X with respect to any attribute subset $B \subseteq C$ are defined as follows:

$$\bar{X}_B^R = \bigcup_{x \in U \wedge [x]_R \cap X \neq \varnothing} [x]_R, \ \underline{X}_B^R = \bigcup_{x \in U \wedge [x]_R \subseteq X} [x]_R.$$

According to the above definitions, we may further define concepts such as the positive region, approximate classification precision and attribute reduction.

In the definition of similarity degree $d(x,y)$ of x and y, K_1, K_2 and K_3 are sets of condition attributes in which the value of x or y is a missing attribute value, the λ_1, λ_2 and λ_3 are their corresponding weight parameters, and their values express the importance respectively. K_4 is a set of condition attributes in which the attribute values of x and y are known and equal. The number of the attributes in K_4 plays a very important role to the similarity degree. Since the value of λ_4 could be as high as possible, λ_4 is constantly 1.

K_5 is a set of condition attributes in which attribute values of x and y are known but not equal. The number of the attributes in K_5 could be as less as possible. If noise is not taken into consideration, λ_5 is 0, otherwise $0 < \lambda_5 < 1$. Hence, the value of λ_5 expresses the degree of noise. Ziarko proposed a variable precision rough set model (VPRS) [15] because the classical rough set theory does not process data with noise. By setting the value of $\beta \in (0.5,1]$, the VPRS model defines concepts such as positive region and approximate classification precision. The less β is, the smaller the border of the objects subset will be, while the bigger the positive region and the negative region will be. Therefore, the VPRS model takes noise into account in its entirety for CISs. However, in the model proposed here, noise is considered in the similarity degree of any two objects. It is especially careful and rational by contrast.

Is the relation R reflexive, symmetric and transitive? It relates to the special IIS and the six parameters' values. The parameter λ is set to be the minimal similarity degree of the two objects subordinated to the same similarity class, and in existing extended models, $\lambda = 1$, $\lambda_5 = 0$. In fact, when $\lambda = 1$ and $\lambda_i \in \{0,1\}$, $i = 1,2,3,4,5$, the relation R is an existing extended model. For example, when $\lambda_4 = 1$ and $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_5 = 0$, R is an indiscernibility relation of the classical rough set theory; when $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 1$ and $\lambda_5 = 0$, R is a tolerance relation; when $\lambda_1 = \lambda_2 = \lambda_4 = 1$ and $\lambda_3 = \lambda_5 = 0$, R is a non-symmetric similarity relation. Thus, the relation R is defined as a generalization relation of the existing extended relations.

For an IIS, the acquired knowledge using the existing deferent extended models might be different. In our model, if the parameters λ_i, $i = 1,2,3,4,5$, take different values, different existing extended models will be generated. Therefore, we may analyze the data using probability statistical methods and set these parameters using a data-driven method without any prior knowledge. It will be our future work.

5 Problems for further study of rough set theory in IISs

In IISs, much uncertain information cannot be estimated exactly. Considering that data reparation might change the information of the original IISs, the extension of the classical rough set theory can be further studied. Taking all the research results into consideration, further study should consider the following issues:

1) For the four types of extended models missing in Table 2, we proposed definitions of possible extended relations. These relations still need testing in applications.

2) By analyzing the definitions of existing extended models, a novel general extended model that needs testing in applications is proposed.

3) Similar to CISs, the rules acquired from IISs can also be divided into certain rules and uncertain rules. Since there is more uncertain information in IISs, the measure of the rules acquired from IISs is very difficult. This is a problem for further studies in the future.

4) In an information system, the attributes are generally considered to be independent variables with the same importance degree. In fact, the importance degree of each attribute is different. Therefore, under the condition that the importance degree of each attribute is different and the attributes are dependent variables, it is an issue yet to be studied to extend the rough set theory.

In conclusion, it is difficult to acquire knowledge from IISs. It is not realistic to acquire exact knowledge from IISs. Thus, how to generate uncertain knowledge from IISs will be a problem for further study.

6 Conclusions

In data mining, many information systems are incomplete. Therefore, how to extract knowledge from IISs becomes a key question.

There are two strategies for processing IISs at present: data reparation and model extension. Data reparation changes the original information, and finally makes the knowledge change in the knowledge acquisition process. Model extension, which is implemented by extending the indiscernibility relation of the classical rough set theory to the inequivalence relation, can acquire knowledge from IISs directly without changing the original information and keeping knowledge unchanged in the knowledge acquisition process. Therefore, the strategy of model extension is found to be suitable for processing IISs. All the extended models can be divided into eight types. In this paper, we analyzed the existing extended models, focusing on those shown in Table 2. The most important task of the application of rough set theory in processing IISs is to construct extended models of the classical rough set theory. Only if the extended model is suitable will the knowledge generated be worthy. This is also our focus for further work in the future.

Acknowledgements This paper was supported by the National Natural Science Foundation of China (Grant Nos. 60573068, 60773113), the Program for New Century Excellent Talents in University (NCET), and the Natural Science Foundation of Chongqing of China (No. 2008BA2017).

References

1. Pawlak Z. Rough sets. International Journal of Computer and Information Sciences, 1982, 11: 341–356
2. Pawlak Z. Rough sets: a new approach to vagueness. Fuzzy Logic for the Management of Uncertainty. New York: John Wiley & Sons, 1992: 105–118
3. Stefanowski J, Tsoukiàs A. On the extension of rough sets under incomplete information. In: Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing (RSFDGrC'99). Lecture Notes In Computer Science, Berlin: Springer, 1999, 1711: 73–82
4. Wang G Y. Rough Set Theory and Knowledge Acquisition. Xi'an: Xi'an Jiaotong University Press, 2001, 38–39 (in Chinese)
5. Grzymala-Busse J W. Rough set strategies to data with missing attribute values. In: Proceedings of the Workshop on Foundations and New Directions in Data Mining, associated with the 3rd IEEE International Conference on Data Mining, 2003, 56–63
6. Kryszkiewicz M. Probabilistic approach to association rules in incomplete databases. In: Proceedings of the First

International Conference on Web-Age Information Management (WAIM'2000). Lecture Notes in Computer Science, Berlin: Springer, 2000, 1846: 133–138
7. Grzymala-Busse J W, Hu M. A comparison of several approaches to missing attribute values in data mining. In: Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing (RSCTC'2000). Lecture Notes in Computer Science, Berlin: Springer, 2001, 2005: 378–385
8. Kryszkiewicz M. Rough set approach to incomplete information systems. Information Sciences, 1998, 112(1): 39–49
9. Stefanowski J, Tsoukiàs A. Incomplete information tables and rough classification. Computational Intelligence, 2001, 17(3): 545–566
10. Stefanowski J, Tsoukiàs A. Valued tolerance and decision rules. In: Proceedings of the Second International Conference on Rough Sets and Current Trends in Computing (RSCTC'2000). Lecture Notes in Computer Science, Berlin: Springer, 2001, 2005: 212–219
11. Wang G Y. Extension of rough set under incomplete information systems. Journal of Computer Research and Development, 2002, 39(10): 1238–1243 (in Chinese)
12. Grzymala-Busse J W. LERS: a system for learning from examples based on rough sets. In: Slowinski R ed. Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publishers, 1992, 3–18
13. Grzymala-Busse J W. A rough set approach to data with missing attribute values. In: Proceedings of the First International Conference of Rough Sets and Knowledge Technology (RSKT'2006). Lecture Notes in Computer Science, Berlin: Springer, 2006, 4062: 58–67
14. Wang G Y. Domain-oriented data-driven data mining based on rough sets. In: Proceedings of International Forum on Theory of GrC from Rough Set Perspective (IFTGrCRSP2006), 2006, 46–46
15. Ziarko W. Variable precision rough sets model. Journal of Computer and System Sciences, 1993, 46(1): 39–59