THE INCOMPATIBLE KNOWLEDGE ELIMINATION IN KNOWLEDGE-INTEGRATION*

Ping Guo, Lian Ye, Li Fan
School of Computer Science, Chongqing University, Chongqing, 400044, P.R.China
E-MAIL: guoping@cqu.edu.cn

Abstract: Knowledge base is the foundation of intelligent system. It is very important to insure the consistency and non-redundancy of knowledge in knowledge base. The redundant, inclusive and incompatible knowledge must be processed in knowledge-integration due to variety of knowledge source. In this paper, we research the incompatible knowledge elimination approach in knowledge-integration based on rough set theory, and present a new knowledge-integration framework, which is effective to improve the efficiency of knowledge-integration.

Key words: rough set,knowledge base,knowledge integration framework

1. INTRODUCTION

In expert system and intelligent information system, the knowledge of knowledge base must be consistent, complete and unambiguous. Due to the difficulty of knowledge-acquisition, the knowledge come from different experts and exterior environment in the domain, so, it may have redundant, inclusive and incompatible knowledge. The purpose of knowledge-integration is to eliminate the redundancy, inclusion and contradiction in knowledge base.

Gaines and Shaw[1], Baral, Kraus and Minker [2] research integration technology of multiple knowledge, and mention that the knowledge-

* This work was supported by National Natural Science Foundation of China, NO. 50378093.

integration plays an important role in building expert system. Medsker, Tan and Turban[3] mention the benefits for developing an expert system by integrating multiple knowledge sources, (1) knowledge reuse, (2) knowledge is more valid, (3) knowledge is more comprehensive, (4) knowledge integration by computers facilitates faster and more inexpensive than that by manpower [4].

Wang, hong and Tesng[5,6,7] propose the knowledge-integration strategies based on genetic algorithm (GA) at the rule-set level in distributed-knowledge environment. The proposed knowledge-integration consists of two phases: knowledge encoding and integration. First, the knowledge is transformed into centralized inner representation and encoded as a fixed-length bit string. Second, the encoding set as initial population is taken to genetic operation. The integrated knowledge base can be obtained by decoding the population of last got.

In the paper, we research the incompatible knowledge elimination approach in knowledge-integration based on rough set theory. The remainder of this paper is organized as follows: stipulate the centralized representation patterns based on whole dictionary in section 2. In section 3, the knowledge set is divided into non-incompatible knowledge set and incompatible knowledge set based on RS classification method. In section 4, the contradiction elimination approach and strategy in incompatible knowledge set is discussed, and that the knowledge set obtained is non-contradiction is proved. A knowledge-integration framework to eliminate incompatible knowledge is presented in section 5. Conclusions are given in Section 6.

2. STIPULATION OF KNOWLEDGE REPRESENTATION

The knowledge from different knowledge sources has different representations. The stipulations for knowledge-integration are as follows:

(1) The whole dictionary is constructed. Knowledge is expressed by using consistent logic predication, variable and logic constant.

(2) All knowledge is denoted as form of production rule

$$r: p \to q \tag{1}$$

Logic precondition p and logic conclusion q are conjunction normal form.

After the vocabularies being centralized encoded, the knowledge in formula (1) is denoted as following pattern:

$$r': p_1 \wedge p_2 \wedge \ldots \wedge p_n \to q$$

Request: the all precondition of r' comprise the same amount of items (when cannot satisfy the request, add several code with value of true), and conclusion of r' is single predication or logic constant.

The integrated knowledge set is denote as $U=\{ u_i \}$, u_i is denoted as:

$$u_i: p_{i1} \wedge p_{i2} \wedge \ldots \wedge p_{in} \rightarrow q_i \quad (2)$$

3. CLASSIFICATION OF KNOWLEDGE SET

Rough set is a common mathematical tool dealing with uncertain problem and uncertain knowledge. R in knowledge set U is constructed as follows:

R: $u_i R u_j \Leftrightarrow$ logic precondition of u_i= logic precondition of u_j, $\forall u_i, u_j \in U$
This is the relation R exists between u_i and u_j iff logic precondition of u_i and u_j are sameness.

That R is equivalence relation in U can be proved. The equivalence class U_R to U can be obtained by R. Each equivalence class is called a concept, viz. an item of knowledge (rule). Different elements belonging to one equivalence class cannot be distinguished to this concept.

The set of knowledge conclusion in U is denoted as Q,

$$Q=\{q_i | u_i \in U,\ q_i \text{ is logic conclusion of } u_i\}$$

The subset U_i of U can be defined by logic conclusion of knowledge:

$$U_i=U(u_i)=\{u_j | u_j \in U,\ \text{logic conclusion of } u_j,\ q_j = \text{logic conclusion of } u_i,\ q_i \}$$

Definition 1. For any subset U_i of U,

$$\underline{Apr}_R(U_i)=\{u | u \in x \ \&\ x \in U_R \ \&\ x \subset U_i\}$$

$$\overline{Apr}_R(U_i)=\{u | u \in x \ \&\ x \in U_R \ \&\ x \cap U_i \neq \emptyset\}$$

$\underline{Apr}_R(U_i)$ and $\overline{Apr}_R(U_i)$ are called the lower-approximation set and the upper-approximation set of U_i respectively.

Definition 2. For any subset U_i of U,

$$BND_R(U_i)=\overline{Apr}_R(U_i)-\underline{Apr}_R(U_i)$$

Is called the boundary set of U_i.

Definition 3. For $u_i, u_j \in U$, if they have the same logic precondition and different logic conclusions, then they are called contradiction.

By lower-approximation set and boundary set of U_i, there is:

Theorem 1. The knowledge in $\underline{Apr}_R(U_i)$ is non-contradiction, and the knowledge in $\underline{Apr}_R(U) = \bigcup_i \underline{Apr}_R(U_i)$ is non-contradiction.

Proof:

First, we prove the knowledge in $\underline{Apr}_R(U_i)$ is non-contradiction.

For $\forall u_j, u_k \in \underline{Apr}_R(U_i)$, by definition 1, $u_j, u_k \in U_i$, so the logic conclusion of u_j and u_k is the same. There are two cases:

(1) When u_j and u_k belong to the same equivalence class of U_R, they have the same logic precondition, so non-contradiction.

(2) When u_j and u_k belong to different equivalence classes of U_R, they have different logic preconditions, so non-contradiction.
Therefore, the knowledge in $\underline{Apr}_R(U_l)$ is non-contradiction.

Second, for $\underline{Apr}_R(U)$, we only need to prove $\forall u_j \in \underline{Apr}_R(U_l)$ and $\forall u_k \in \underline{Apr}_R(U_m)$($U_l \neq U_m$)is non-contradiction. Due to $U_l \neq U_m$, viz. u_j and u_k have different logic conclusions. According to the definition 1, $\exists x$, $y \in U_R$, and $x \subset U_l$, $y \subset U_m$, so $u_j \in x$, $u_k \in y$ and $x \neq y$, otherwise $u_j \in x = y \subset U_m$ is incompatible with different logic conclusions of u_j and u_k. Therefore, u_j and u_k have different logic preconditions.

Summarize above, u_j and u_k not only have different logic preconditions, but also have different logic conclusions. So they are non-contradiction.

[♦]

Theorem 2. If $|BND_R(U_l)|>1$, the knowledge in $BND_R(U_i)$ has contradiction. $BND_R(U)=\bigcup_i BND_R(U_l)$. If $|BND_R(U)|>1$, the knowledge in $BND_R(U)$ has contradiction.

Proof:

First, we prove that the knowledge in $BND_R(U_l)$ has contradiction when $|BND_R(U_i)|>1$.

By definition 2, $\exists x \in U_R$, so as to $x \cap U_i \neq \emptyset$ and $x \not\subset U_i$, viz.$\exists u_j$, $u_k \in x$, so that $u_j \in U_i$, $u_k \notin U$. So u_j and u_k have the same logic precondition and different logic conclusions, u_j and u_k are contradiction. Viz. the knowledge in $BND_R(U_i)$ has contradiction.

Second, for $BND_R(U)$, only need to prove that there exists i so as to $|BND_R(U_i)|>1$ when $|BND_R(U)|>1$.

By $|BND_R(U)|>1$, there exists m so as to $|BND_R(U_m)|>0$. By the definition of $BND_R(U_m)$ to know: $\exists x \in U_R$, so that $x \cap U_m \neq \emptyset$ and $x \not\subset U_m$. Apparently, there exists $|x|>1$, viz. $|BND_R(U_m)|>1$.

Summarize above, the knowledge in $BND_R(U)$ has contradiction.

[♦]

We can prove, $\underline{Apr}_R(U) \cap BND_R(U)=\emptyset$. So U can be divided up:

$$U=\underline{Apr}_R(U) \cup BND_R(U) \tag{3}$$

By theorem 1 and theorem 2, the $BND_R(U)$ only need be processed to eliminate the incompatible knowledge in U.

4. INCOMPATIBLE KNOWLEDGE PROCESS IN BOUNDARY SET

By formula (3), it needs to eliminate the incompatible knowledge in boundary set $BND_R(U)$ to make non-contradiction of knowledge in U. There are several strategies for process [5,6,7,8]. A test dataset is introduced to eliminate the contradiction of knowledge in $BND_R(U)$ in this paper.

Definition 4. Assume the test object set Ω. That object $o \in \Omega$ is correctly predicted by knowledge u_i means the logic precondition of u_i is satisfied with o and logic conclusion of u_i is consistent with conclusion of o. Denote the object set correctly predicted by u_i as Ω_{u_i}. That object o is wrongly predicted by knowledge u_i means the logic precondition of u_i is satisfied with o but logic conclusion of u_i is inconsistent with conclusion of o. Denote the object set wrongly predicted by u_i as $\not\Omega_{u_i}$.

Using test object set Ω, calculate the object set Ω_{u_i} and $\not\Omega_{u_i}$ predicted by each item of knowledge u_i in $BND_R(U)$, and calculate the parameters:

$$A(u_i)=\begin{cases} 0 & \text{if } |\Omega_{u_i}|+|\not\Omega_{u_i}|=0 \\ \dfrac{|\Omega_{u_i}|}{|\Omega_{u_i}|+|\not\Omega_{u_i}|} & \text{if } |\Omega_{u_i}|+|\not\Omega_{u_i}|\neq 0 \end{cases} \quad (4)$$

$$E(u_i)=\sum_{o\in\Omega}\frac{\Phi(u_i,o)}{\sum_{u\in BND_R(U)}\Phi(u,o)} \quad (5)$$

Thereinto:

$$\Phi(u_j,e)=\begin{cases} \mathbf{1} & \text{if o is correctly predicted by } u_j \\ \mathbf{0} & \text{otherwise} \end{cases}$$

$$C(u_i)=\frac{|\Omega_{u_i}|+|\not\Omega_{u_i}|}{|\Omega|} \quad (6)$$

The calculative result of formulas (4), (5), (6) is called accuracy, utility and coverage respectively. As to every item of knowledge, the ideal situation is the maximum of accuracy, utility and coverage. But these cannot usually be satisfied at the same time. Calculate the following integrative criterion:

$$f(u_i)=A(u_i)*E(u_i)*C(u_i) \quad (7)$$

The strategy for eliminating incompatible knowledge in $BND_R(U)$ by calculative result of formula (7) is: the calculative results of formula (7) are sorted in descending. If u_i and u_j are incompatible knowledge in $BND_R(U)$ and $f(u_i)$ is in front of $f(u_j)$ in order then delete u_j in $BND_R(U)$, else delete u_i.

The result set which is deleted incompatible knowledge in $BND_R(U)$ is denoted as $\underline{BND}_R(U)$, and denote

$$\underline{U}=\underline{Apr}_R(U)\cup\underline{BND}_R(U) \quad (8)$$

Theorem 3. The knowledge in $\underline{U}$ is non-contradiction.

Proof:

For $\forall u_j, u_k \in \underline{U}$, we divide three steps to prove the theorem.

(1) $u_j, u_k \in \underline{Apr}_R(U)$. By theorem 1, u_j is consistent with u_k.

(2) $u_j \in \underline{Apr}(U)$, $u_k \in \underline{BND}_R(U)$. Assume u_j and u_k are incompatible, viz.: u_j and u_k have the same logic precondition, but different logic conclusions.

By definition 1, $\exists m$ and $x \in U_R$, so that: $u_j \in x$, $x \subset U_m$

Due to u_j and u_k having the same precondition and definition of U_R, exist $u_k \in x \subset U_m$. Viz., u_j and u_k have the same conclusion q_m. This is contradicted with assumption of different logic conclusions of u_j and u_k. So u_j and u_k are non-contradiction.

(3) u_j, $u_k \in \underline{BND}(U)$. u_j and u_k are non-contradiction by the construction procedure of $\underline{BND}(U)$.

Summarize above, the theorem is correct. [♦]

5. KNOWLEDGE-INTEGRATION FRAMEWORK

By the conclusion of above sections, the proposed framework for eliminating incompatible knowledge in knowledge-integration is shown in figure1. Decoding procedure is the inverse procedure of knowledge encode. The knowledge in $\underline{Apr}_R(U) \cup \underline{BND}_R(U)$ can be transformed into knowledge representation comprehensible by whole dictionary. The additive value of true in logic precondition of knowledge is deleted to predigest the knowledge representation.

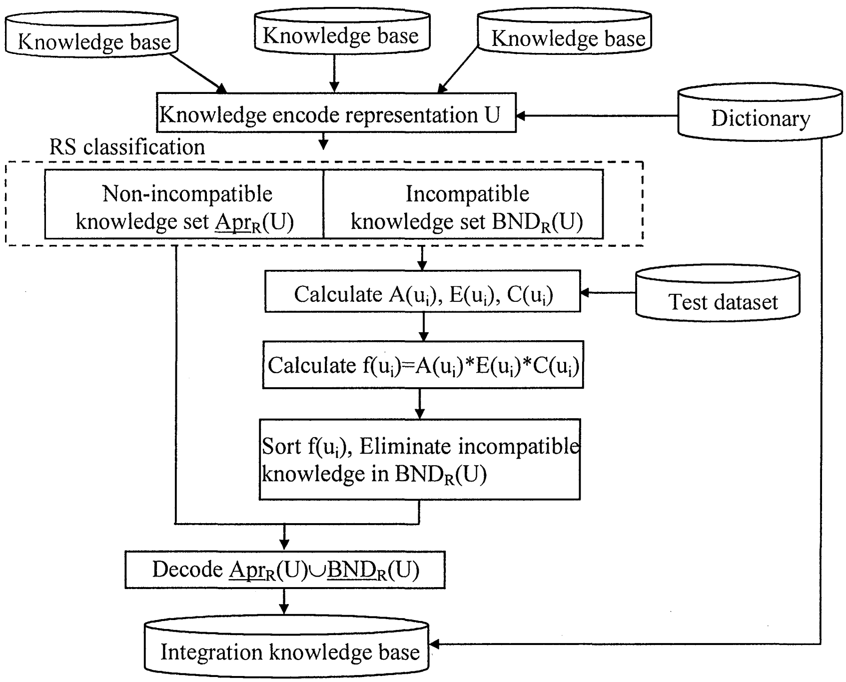

Figure 1. Knowledge-integration framework

6. CONCLUSION

It is one of pivotal problems in knowledge-integration research to eliminate the incompatible knowledge in knowledge base. In this paper, the knowledge in knowledge base is divided into two types by RS theory: non-incompatible knowledge set and incompatible knowledge set. To each item of knowledge in incompatible knowledge set, calculate the accuracy, utility and coverage of test data, provide a feasible approach to eliminate incompatible knowledge, and prove that the knowledge set acquired is non-contradiction. At the same tine, we present a framework to eliminate incompatible knowledge in knowledge base based on the theory.

REFERENCES

[1] Gaines, B. R., & Shaw, M. L. (1993). Eliciting knowledge and transferring it effectively to a knowledge-based system. IEEE Transaction on Knowledge and Data Engineering, 5 (1), 4–14.

[2] Baral, C., Kraus, S., & Minker, J. (1991). Combining multiple knowledge bases. IEEE Transactions on Knowledge and Data Engineering, 3 (2), 208–220.

[3] Yuan, Y., & Zhuang, H. (1996). A genetic algorithm for generating fuzzy classification rules. Fuzzy Sets and Systems, 84, 1–19.

[4] Medsker, L., Tan, M., & Turban, E. (1995). Knowledge acquisition from multiple experts: problems and issues. Expert Systems with Applica-tions, 9 (1), 35–40.

[5] Wang, C. H., Hong, T. P., & Tseng, S. S. (1997). Knowledge integration by genetic algorithms. Proceedings of the Seventh International Fuzzy Systems Association World Congress, 2, 404–408.

[6] Wang, C. H., Hong, T. P., & Tseng, S. S. (1998). A genetic fuzzy-knowledge integration framework. The Seventh International Conference of Fuzzy Systems, 1194–1199.

[7] Wang, C. H., Hong, T. P., & Tseng, S. S. (2000). Integrating membership functions and fuzzy rule sets from multiple knowledge sources, Fuzzy Sets and Systems, 112, 141–154.

[8] Wang, C. H., Hong, T. P., & Tseng, S. S. (2000). A Genetics-Based Approach to Knowledge Integration and Refinement. Journal of Information Science and Engineering, 17, 85-94.

[9] Mathias,K.E.&Whity,L.D.(1994),Transforming the Search Spacs with Gray Coding, Proc. of the 1st IEEE Intl. Conf. On Evolutionary Computation,Orlando,Florid,USA,IEEE Press,519-542.

[10] Wang, C. H., Hong, T. P., & Tseng, S. S. (2000). A Coverage-based Genetic Knowledge-integration strategy, Experty Systems with Applications,19(2000),9-17.