

Research on the Approach of Knowledge Acquisition in Expert Systems based on Rough Sets Theory

Yehong HAN

Institute of Information Science and Engineering, Qilu Normal University, Jinan, 250014, China

email: sdzzyh@163.com

Keywords: Expert Systems; Rough Sets; Knowledge Discovery; Data Mining; Artificial Intelligence

Abstract. The research on expert systems has been a hotspot. In order to resolve the problems about knowledge acquisition in complete and incomplete knowledge representation systems, the algorithm of knowledge acquisition based on rough set theory is proposed. Let the knowledge representation system be a decision table, the useful knowledge can be obtained by simplification of the decision table. The algorithm of computation of condition attributes' reductions and the algorithm of computation of reduction of decision rule are researched to simplify the decision table. When there are missing values in knowledge representation systems, the missing data should be completed by the algorithm of missing values' completion firstly. This algorithm of knowledge acquisition based on rough set theory can obtain a minimal set of decision rules which can be used to reason in expert systems. The approach of knowledge acquisition can be used to simplify uncertain and incomplete knowledge representation systems.

Introduction

With the development of the application system of the artificial intelligence and the knowledge engineering, the expert system [1] has been a hot area. It was successfully applied in many domains. With the enhancement of demand, many questions such as the knowledge acquisition [2,3] and the uncertain knowledge reasoning are exposed.

Rough sets theory [4] was proposed by Z. Pawlak in 1982. The theory provides a new tool for dealing with classification of fuzzy, inaccurate or incomplete information. The main idea of the theory is that rules of decision and classification can be obtained by reduction of knowledge remaining the classification ability of information system unchanged.

In this paper, a model of an expert system based on rough sets theory as shown in fig. 1 is given, and the knowledge acquisition mechanism based on the model will be researched in detail. An improved algorithm of knowledge acquisition from complete knowledge representation system based on rough sets theory and an improved algorithm of missing values' completion in incomplete knowledge representation system will be proposed.

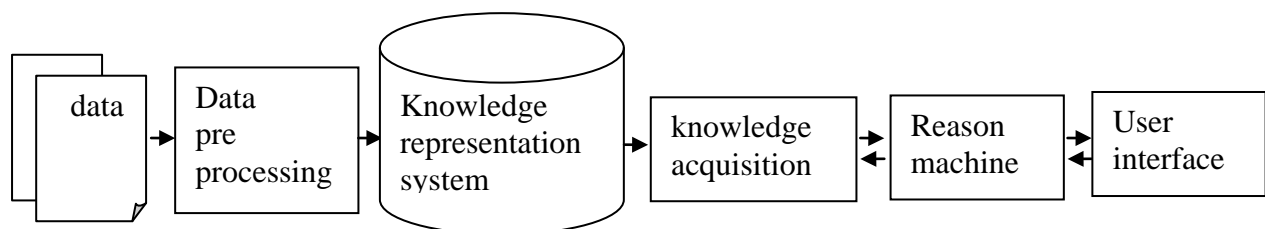


Fig.1 a model of expert systems based on rough sets theory

Knowledge representation system [5] can be formulated as a pair $S = (U, A)$. U is a nonempty and finite set called the universe. A is a nonempty and finite set of primitive attributes. Every primitive attribute $a \in A$ is a function $a: U \rightarrow V_a$ (V_a is the set of values of a , called the domain of a). If $u \in U$, $a \in A$ and $v_a \in V_a$, the knowledge representation system will be called complete knowledge representation system, otherwise the knowledge representation system is a incomplete

knowledge representation system. Let $C, D \subset A$ (where C, D called condition and decision attributes respectively), the knowledge representation system with distinguished condition and decision attributes will be called a decision table, and will be denoted $S = (U, A, C, D)$.

Knowledge Acquisition from Complete Knowledge Representation Systems Based on Rough Sets Theory

The process of knowledge acquisition in expert systems based on rough sets theory is actually simplification of knowledge representation systems [6]. Simplification of Knowledge representation system is simplified as follow.

Let $S = (U, A, C, D)$ be a knowledge representation system. If $POS_{C-\{a\}}(D) = POS_C(D)$, the attribute a is dispensable in knowledge representation system, otherwise the attribute a is indispensable. If all attributes $a \in C$ are indispensable, the knowledge representation system is independent. Let $R \subseteq C$. If $POS_R(D) = POS_C(D)$ and the decision table $S' = (U, R \cup D, R, D)$ is independent, R which can be denoted $RED(C, D)$ is a reduction of condition attributes in the knowledge representation system, S' is a reduction of the decision table $S = (U, A, C, D)$.

The set of all indispensable attributes in condition attributes will be called the core of the decision table, and will be denoted $CORE(C, D)$. $CORE(C, D) = \bigcap RED(C, D)$.

Let $\theta \rightarrow \psi$ be a RD -rule and $a \in R$, If $\|\theta \wedge \bar{\psi}\|_S = \|\theta / (R - a) \wedge \bar{\psi}\|_S$, the attribute a is dispensable in the rule $\theta \rightarrow \psi$, otherwise the attribute a is indispensable. If all attributes $a \in R$ are indispensable, the rule $\theta \rightarrow \psi$ is independent. If $R' \subseteq R$, $\theta / R' \rightarrow \psi$ and $\|\theta \wedge \bar{\psi}\|_S = \|\theta / R' \wedge \bar{\psi}\|_S$, $\theta / R' \rightarrow \psi$ is a reduction of RD -rule: $\theta \rightarrow \psi$.

The set of all indispensable attributes of RD -rule: $\theta \rightarrow \psi$ will be called the core of the rule and will be denoted $CORE(\theta \rightarrow \psi)$.

The approach to simplification of knowledge representation systems [7] presented in this paper consists of the following steps:

- (1) Computation of condition attributes' reductions;
- (2) Elimination of duplicate rows in the decision table $S' = (U, RED(C, D) \cup D, RED(C, D), D)$ (remark: rows do not represent here description of any real objects);
- (3) Computation of reduction of every rule in the knowledge representation system.

The Approach to Simplify Knowledge Representation Systems

- (1) The algorithm of computation of condition attributes' reductions

Computation of condition attributes' reductions [8,9] has been proved to be a Non-deterministic Polynomial Complete problem. In order to improve the efficiency of knowledge acquisition, a lot of researches on approaches to compute reductions of condition attributes have focused on the approximate reduction of attributes. In this paper, a algorithm of approximate reduction based on the importance of attributes will be proposed.

Let $S = (U, A, C, D)$ be a decision table, $R \subseteq C$, and $a \in C - R$, the importance of attributes a which will be denoted $importance(a, R, D)$ can be formulated as follows:

$$importance(a, R, D) = \gamma_{R \cup \{a\}} - \gamma_R = \frac{|POS_{R \cup \{a\}}(D)|}{|U|} - \frac{|POS_R(D)|}{|U|} \quad (1)$$

The algorithm of approximate reduction based on the importance of attributes can be described as follows:

Input: a knowledge representation system $S = (U, A, C, D)$, a set of important attributes X ;

Output: a set of attribute reduction R

Algorithm:

$$R = X ;$$

```

IF  $R = \phi$  THEN
  FOR each  $a \in C$ 
    {compute  $|POS_{\{a\}}(D)|$ ;
     find the attribute  $a$  which meets the value of  $|POS_{\{a\}}(D)|$  is biggest;
    }
   $R = \{a\}$ ;
While  $|POS_R(D)| \neq |POS_C(D)|$  DO
  { FOR each  $a \in C - R$ 
    {compute  $|POS_{R \cup \{a\}}(D)|$ ;
     find the attribute  $a$  which meets the value of  $|POS_{R \cup \{a\}}(D)|$  is biggest;
    }
     $R = R \cup \{a\}$ ;
  }
FOR each  $a \in R - Y$ 
  {  $R = R - \{a\}$ ;
    IF  $|POS_R(D)| \neq |POS_C(D)|$  THEN
       $R = R \cup \{a\}$ ;
  }

```

(2) The algorithm of computation of reduction of decision rule.

Let $S = (U, A, C, D)$ be a decision table, the support of rule $\theta \rightarrow \psi$ in the decision table S can be denoted $\text{sup port}(\theta \rightarrow \psi)$, and will be defined as formula (2).

$$\text{sup port}(\theta \rightarrow \psi) = \text{card}|\theta \rightarrow \psi|_S \quad (2)$$

Accordingly, the support of formula θ in the decision table S can be denoted $\text{sup port}(\theta)$, and will be defined as formula (3).

$$\text{sup port}(\theta) = \text{card}|\theta|_S \quad (3)$$

Obviously, if $\text{sup port}(\theta) \leq \min \text{sup}$, $\text{sup port}(\theta \wedge \varphi) \leq \min \text{sup}$ ($\min \text{sup}$ is a value given by person)[10].

This is a important proposition. The following algorithm can be obtained according to the proposition.

Input: the decision table $S' = (U, \text{RED}(C, D) \cup D, \text{RED}(C, D), D)$, $\min \text{sup}$.

Output: a minimal set of decision rules RUL .

Algorithm: For ($k = 1$; $k = |R|$; $k++$) do

```

{  $C_k = \phi$ ;
   $R_k = \phi$ ;
}

```

Let C_1 be the set of rules which only have a condition attribute in S'

For ($k=1$; $k=|R|$; $k++$)

{For $\forall \theta \rightarrow \psi \in C_k$

{compute $\sup port(\theta)$ and $\sup port(\theta \rightarrow \psi)$;

If $\sup port(\theta \rightarrow \psi) \leq \min \sup$ then

$C_k = C_k - \{\theta \rightarrow \psi\}$;

Else if $\sup port(\theta \rightarrow \psi) = \sup port(\theta)$ Then

{ $C_k = C_k - \{\theta \rightarrow \psi\}$;

$R_k = R_k \cup \{\theta \rightarrow \psi\}$;

}

}

For $\forall \theta \rightarrow \psi \in C_k$ and $\theta' \rightarrow \psi' \in C_k$

{ $F = S_CA(\theta \rightarrow \psi) \cap S_CA(\theta' \rightarrow \psi')$;

If ($\psi = \psi'$ and $|F| = k-1$ and $\theta / F = \theta' / F$) Then

$C_{k+1} = C_{k+1} \cup \theta / F \wedge \theta / (S_CA(\theta \rightarrow \psi) - F) \wedge \theta' / (S_CA(\theta' \rightarrow \psi') - F) \rightarrow \psi$;

}

For $\forall \theta \rightarrow \psi \in C_{k+1}$

If ($\|\theta \wedge \psi\|_S = \phi$) Then $C_{k+1} = C_{k+1} - \{\theta \rightarrow \psi\}$;

}

$RUL = \bigcup_k R_k$;

For $\forall \theta \rightarrow \psi \in RUL$ and $\theta' \rightarrow \psi \in RUL$

If ($\theta' \rightarrow \theta$) Then $RUL = RUL - \{\theta' \rightarrow \psi\}$;

Knowledge Acquisition from Incomplete Knowledge Representation Systems Based on Rough Sets Theory

Knowledge acquisition system is to obtain useful knowledge from a large number of historical data. But there are many missing data in incomplete knowledge representation systems. In order to obtain useful knowledge, completion of missing values should be researched firstly. Algorithm of missing values' completion [11] include mean completer, combinatorial completer, ROUSTIDA, and so on. In this paper, an improved algorithm of ROUSTIDA will be introduced.

Let $S = (U, A, C, D)$ be a decision table, $U = \{x_1, x_2, \dots, x_n\}$, $C = \{c_i | i = 1, \dots, m\}$, the discernable matrix can be denoted M , will be defined as follows:

$$M(i, j) = \begin{cases} c_k | c_k \in C \wedge c_k(x_i) \neq c_k(x_j) \wedge c_k(x_i) \neq * \wedge c_k(x_j) \neq *, i \neq j; \\ 0, i = j. \end{cases} \quad (4)$$

Where $i = 1, \dots, m$, $j = 1, \dots, n$, $*$ represent a missing value.

Let $S = (U, A, C, D)$ be a decision table, $U = \{x_1, x_2, \dots, x_n\}$, $C = \{c_i | i = 1, \dots, m\}$, the set of missing attributes of object x_i MAS_i , the set of indistinctive objects of object x_i NS_i , and the set of missing objects of knowledge representation system S can be defined as formula (5), formula (6) and formula(7) respectively.

$$MAS_i = \{c_k | c_k(x_i) = *, k = 1, \dots, m\} \quad (5)$$

$$NS_i = \{j \mid M(i, j) = \Phi, i \neq j, j = 1, \dots, n\} \quad (6)$$

$$MOS = \{i \mid MAS_i \neq \Phi, i = 1, \dots, n\} \quad (7)$$

The algorithm of missing values' completion will be shown as follows.

Input: incomplete knowledge representation system S' ;

Output: complete knowledge representation system S ;

Algorithm: compute M^0, MAS^0, MOS^0 ;

```

r=0;
FOR  $\forall i \in MOS^r$ 
    { compute  $NS_i^r$ ;
      Compute  $card(NS_i^r)$ ;
    }
FOR  $\forall i \in MOS^r$  AND  $\forall k \in (i \mid i \in \{1, \dots, m\} \wedge c_i \in MAS_i^r)$ 
    { IF  $card(NS_i^r) = 0$  THEN  $c_k(x_i^{r+1}) = *$ ;
      IF  $card(NS_i^r) = 1$  THEN
          {  $j \in NS_i^r$ ;
            IF  $c_k(x_i^r) = *$  THEN  $c_k(x_i^{r+1}) = *$ 
            ELSE  $c_k(x_i^{r+1}) = c_k(x_i^r)$ ;
          }
      IF  $card(NS_i^r) > 1$  THEN
          IF  $\exists j_0, j_1 \in NS_i^r \Rightarrow c_k(x_{j_1}^r) \neq c_k(x_{j_0}^r)$  THEN  $c_k(x_i^{r+1}) = *$ 
          ELSEIF  $\exists j_0 \in NS_i^r \Rightarrow c_k(x_{j_0}^r) \neq *$  THEN  $c_k(x_i^{r+1}) = c_k(x_i^r)$ 
          ELSE  $c_k(x_i^{r+1}) = *$ 
      compute  $M^{r+1}, MAS^{r+1}, MOS^{r+1}$ ;
       $r = r + 1$ ;
    }

```

Partial missing data can be completed by the algorithm. If the knowledge representation system is incomplete, missing data should be completed by other method.

When the missing data is completed, useful knowledge can be acquitted by algorithm of knowledge acquisition from complete knowledge representation system based on rough sets theory proposed above.

Conclusion

In this paper, a basic model structure of an expert system based on rough set theory is given by improving the traditional model structure, and the knowledge acquisition algorithm based on the model are researched in detail. In order to obtain useful knowledge by simplification of knowledge representation systems, the algorithm of computation of condition attributes' reductions and the algorithm of computation of reduction of decision rule are introduced. An algorithm of knowledge acquisition from complete knowledge representation system based on rough sets theory and an improved algorithm of missing values' completion in incomplete knowledge representation system are proposed. The approach of knowledge acquisition in expert systems based on rough sets theory proposed in this paper can be used to simplify both complete and incomplete knowledge representation systems, but it is an algorithm of approximate reduction, there is a low accuracy in incomplete knowledge representation systems. The algorithm must be researched forward.

Acknowledgement

This work was supported by the Natural Science Foundation of Shandong Province, China (Grant No. ZR2013FL010).

References

- [1] Wright J R, Wiggins L L, Jain R K, et al. Expert systems in environmental planning[M]. Springer Publishing Company, Incorporated, 2011.
- [2] Lee C S, Wang M H. A fuzzy expert system for diabetes decision support application[J]. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 2011, 41(1) 139-153.
- [3] Fasanghari M, Montazer G A. Design and implementation of fuzzy expert system for Tehran Stock Exchange portfolio recommendation[J]. Expert Systems with Applications, 2010, 37(9) 6138-6147.
- [4] Z. Pawlak. Rough sets and intelligent data analysis. Information Sciences, 2002, 147 1-12.
- [5] Zhang J, Li T, Ruan D, et al. Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems[J]. International Journal of Approximate Reasoning, 2012(4) 620-635.
- [6] Słowiński R, Greco S, Matarazzo B. Rough-set-based decision support[M]. Search Methodologies. Springer US, 2014: 557-609.
- [7] Qian Y, Liang J, Pedrycz W, et al. Positive approximation: an accelerator for attribute reduction in rough set theory[J]. Artificial Intelligence, 2010, 174(9) 597-618.
- [8] Artificial Intelligence in Engineering Design: Volume III: Knowledge Acquisition, Commercial Systems, And Integrated Environments[M]. Elsevier, 2012.
- [9] Dymova L, Sevastianov P, Bartosiewicz P. A new approach to the rule-base evidential reasoning: Stock trading expert system application[J]. Expert systems with Applications, 2010, 37(8) 5564-5576.
- [10] Park Y, Park J. An expert system for short term load forecasting by fuzzy decision[C]. Power Systems and Power Plant Control 1989: Selected Papers from the IFAC Symposium, Seoul, Korea, 22-25 August 1989. Elsevier, 2014: 461.
- [11] Qian Y, Liang J, Dang C. Incomplete multigranulation rough set[J]. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 2010, 40(2) 420-431.