2004－ICS－138 (10)
2004／12／7

A Framework For Knowledge Discovery in Multivalued Tables Using Rough Set Approach

NGUYEN THANH THUY † and NGUYEN NGOC HIEU ††

In data mining area, researchers have proposed methods to discover decision rules that are based on rough set approach. Most of these methods do not have ability to handle multivalued information systems, though this type of information systems appears frequently in practical problems[1),6)]. In this paper, we propose an approach based on rough set to handle multivalued information systems. With the model designed from this approach, we propose two algorithms to discover minimal decision rules from multivalued information systems. In addition, we make a comparison between this method and other classical ones.

1. Introduction

The theory of rough set is a recent approach to deal with incomplete information. There are many types of incomplete information, but in this paper, we consider the incomplete information whose vagueness results from multivalued attributes. This kind of incomplete information systems is called multivalued information systems or multivalued tables (MIT). Let us consider an example of MIT in Figure 1.

In this system, **Diagnosis** and **Symptoms** attributes are multivalued ones. **Thrombosis** attributes is assigned classify attribute by the experts. It is very difficult to know which rules experts use for classification is also that is the task we are dealing with.

In the classical rough set approach, two objects are considered as mutually indiscernible if their value pairs are identical on all attributes. When considering MIT, if we replace comparing single values with comparing the sets of values by using the set theory equality relation, we obtain decision rules of which conditional parts have the form $A_j = d_{ij}$, where A_j is j^{th} attribute, d_{ij} is a set of i^{th} object's atomic values of attribute A_j. Such conditions may be too strict. The reason is that very few objects are absolutely equal on every attribute-value; and some values in d_{ij} may be unnecessary for classification purposes. Therefore, using sets of values in decision rules leads to increase the complexity as well as limits the induction power.

A method to discover knowledge from MIT by transforming MIT into single-valued information systems is proposed in[2)]. Each object of original system is transformed into a set of objects in the new system. In this case, a wrong possibility to distinguish the objects or an incorrect classification can be created (one object can belong to different classes).

In this paper, we propose an approach to handle MIT by replacing equality relation in classical rough set approach with a certain equivalence relation on the domain of each attribute. Hence equivalent objects can be allowed to have different attribute-values. Moreover, decision rules describe the relation of attribute-value classes (instead of describing relation of attribute-values in the way of classical decision rules[10)]).

This paper consists of four parts: 1) Introduction 2) Basic notions and properties 3) Discovery of decision rules 4) Conclusions.

2. Basic Notions and Properties

2.1 Multivalued Information Systems

Definition 1: Let $MIT = (OB, AT, \rho)$, in which $OB = \{O_1, O_2, \ldots, O_n\}$ is a set of objects, $AT = \{A_1, A_2, \ldots, A_m\}$ is a set of attributes. Let D_j is domain of A_j, each D_j has an equivalence relation $\rho_j \in \rho$ that partitions D_j into equivalent classes $D_j/\rho_j = \{[a_i]_{\rho_j} | a_i \in D_j\}$.

Each $O_i \in OB$ has the following form:

$O_i = (d_{i1}, d_{i2}, \ldots, d_{im})$

where $d_{ij} = \{a_1, a_2, \ldots, a_k\} \subseteq D_j$. Then O_i can be expressed in the form

† Faculty of Information Technology, Hanoi University of Technology
†† Faculty of Information Technology, Vinh University

ID	aCL IgG	...	Diagnosis	Symptoms	Thrombosis
1124385	0	...	SjS	thrombocytepenia	3
1430761	2	...	SLE	CNS lupus	2
1936676	1150.7	...	abortion	abortion, epilepsy	1
3192610	80.3	...	APS	CVA, epilepsy	1
3296266	1.5	...	SLE, SjS, Cliogloblin+	CNS lupus	2
4061234	4.3	...	SLE	Apo, convulsion	1
4746661	1.1	...	SjS(CNS)	CNS lupus	2
4763954	374.12	...	ITP, APS	thrombocytopenia	3
5512586	3.4	...	APS, Budd-Chiari	Budd-Chiari	1

Fig. 1 A multivalued information system[6)]

$d_{i1}/\rho_1, d_{i2}/\rho_2, \ldots, d_{im}/\rho_m)$, where d_{ij}/ρ_j are all equivalent classes that are created by equivalence relation ρ_j. Notation $d_{ij}\rho_j d_{kj}$ or $A_j(O_i)\rho_j A_j(O_k)$ is used to denote $d_{ij}/\rho_j \equiv d_{kj}/\rho_j$. Assuming $B \subseteq AT$, symbol ρ_B indicates an equivalence relation that is created by all equivalence relations ρ_j, where $A_j \in B$. Then, ρ_B partitions OB into equivalence classes.

$$OB/\rho_B = \{[O_i]_{\rho_B} | O_i \in OB, B \subseteq AT\}.$$

The relation $\rho = \rho_{AT}$ classifies all objects of OB of MIT.

Definition 2: A multivalued decision information system (MDT) is a MIT with structure $MDT = (OB, AT \cup DC, \rho)$, in which AT is a set of condition attributes, DC is a set of decision attributes, $(AT \cap DC = \emptyset)$.

Example 1: Let $MIT = (OB, AT, \rho)$, in which $AT = \{A, B\}$, $D_A = \{a_1, a_2, a_3\}$ is the domain of A. There is an equivalence relation ρ_A on D_A that creates equivalence classes $\{\{a_1, a_3\}, \{a_2\}\}$. $D_B = \{b_1, b_2\}$ is domain of attribute B with ρ_B as an equality relation.

We have: $D_A/\rho_A = \{[a_1], [a_2]\} = \{[a_3], [a_2]\}$; $D_B/\rho_B = \{[b_1], [b_2]\}$. Conspicuously, OB/ρ_{AB} can have at most 9 classes, while in classical approach OB can have at most 21 classes.

So, considering MIT helps to reduce a large number of equivalence classes, simplifying the process of data mining, which result in the number of discovered rules sharply reduced, meanwhile ensuring to cover the set of all objects in the table.

2.2 Approximations of Set

Definition 3: Let $MIT = (OB, AT, \rho)$, $X \subseteq OB$ and $B \subseteq AT$. We defines:

- $\underline{B}_{\rho_B}X$ is *lower approximation* of X by ρ_B if $\underline{B}_{\rho_B}X = \bigcup\{O_i \in OB | [O_i]_{\rho_B} \subseteq X\} = \bigcup\{Y \in OB/\rho_B | Y \subseteq X\}$.
- $\overline{B}_{\rho_B}X$ is *uppper approximation* of X by ρ_B if $\overline{B}_{\rho_B}X = \bigcup\{O_i \in OB | [O_i]_{\rho_B} \cap X \neq \emptyset\} = \bigcup\{Y \in OB/\rho_B | Y \cap X \neq \emptyset\}$

Similar to classical information system, $\underline{B}_{\rho_B}X$ is a set of objects that *certainly* belong to X, while $\overline{B}_{\rho_B}X$ is a set of objects that *possibly* belonging to X. Symbol $BN_{\rho_B}(X) = \overline{B}_{\rho_B}X \backslash \underline{B}_{\rho_B}X$ is *boundary region* of X that consists of all of objects that can not be certainly classified into X in B. $OB \backslash \overline{B}_{\rho_B}X$ is *outside region* of X consisting all of objects that does not certainly belong to X.

X is said to be *rough* if its boundary region is non-empty, otherwise X is crisp.

The properties of approximations in MIT are absolutely alike in classical models. The way to prove these properties is similar to the way in[8)].

Definition 4: Let $MDT = (OB, AT \cup DC, \rho)$. $POS_{AT}(DC)$ is called *AT-positive region* of DC if

$$POS_{AT}(DC) = \bigcup_{X \in OB/\rho_{DC}} \underline{AT}_{\rho_{AT}}(X)$$

2.3 Rough membership function

Definition 5: Let $MIT = (OB, AT, \rho)$, $X \subseteq OB$, $B \subseteq AT$, $x \in OB$.

Rough membership function

$$\mu_X^{\rho_B}(x) = \frac{|[x]_{\rho_B} \cap X|}{|[x]_{\rho_B}|}$$

determines the degree that x belongs to X, if considering only in B. Value of function $\mu_X^{\rho_B}(x)$

can be interpreted as conditional probability. Conspicuously, $\mu_X^{\rho_B} \in [0, 1]$.

Rough membership function can be used to redefine approximations and boundary region of a set as follow:

$$\underline{B}_{\rho_B}X = \{x \in OB | \mu_X^{\rho_B}(x) = 1\}$$
$$\overline{B}_{\rho_B}X = \{x \in OB | \mu_X^{\rho_B}(x) > 0\}$$
$$BN_{\rho_B}(X) = \{x \in OB | 0 < \mu_X^{\rho_B}(x) < 1\}$$

Rough membership function can be generalized as follow:

$\mu_{(X,Y)} = \frac{|X \cap Y|}{|X|}$, in which $X, Y \subseteq OB$, $X \neq \emptyset$, determine degree of relation between X and Y.

2.4 Dependency of attributes

Definition 6: Let $MIT = (OB, AT, \rho)$. Consider two sets of attributes $B, C \subseteq AT$. We say C *is called absolutely dependent on* B, denoted by $B \xrightarrow{\rho} C$ if OB/ρ_B is softer than OB/ρ_C. C is called *dependent on* B *at level* k $(0 \leq k \leq 1)$, denoted by $B \xrightarrow{\rho}_k C$, where $k = \gamma(B, C) = \frac{|POS_B(C)|}{|OB|}$. Conspicuously, $\gamma(B, C) = \sum_{X \in OB/\rho_C} \frac{|\underline{B}_{\rho_B}(X)|}{|OB|}$.

We have a remark that $B \xrightarrow{\rho} C$ if and only if $B \xrightarrow{\rho}_1 C$.

2.5 Reduction of Information system

Considering $MIT = (OB, AT, \rho)$ with n objects. MIT may contain some redundant attributes and attribute-values i.e if removed, they do not affect the classification. Thus we only need to keep indispensable attributes and attribute-values that maintain the equivalence relation and approximations of a set. We call the minimal subset of such attributes is the reduction of an information system.

Definition 7: Let $MIT = (OB, AT, \rho)$. A non-empty subset $B \subseteq AT$ with subsets $V_{A_j} \subseteq D_j/\rho_j$ ($A_j \in B, D_j$ is domain of A_j) is called a reduction of AT, denoted as $Reduct_\rho(AT)$ if and only if $OB/\rho_B \equiv OB/\rho_{AT}$.

Definition 8: Let $MDT = (OB, AT \cup DC, \rho)$. $A \in AT$ is called a *dispensable attribute* in MDT if $POS_{AT}(DC) = POS_{(AT-\{A\})}(DC)$; otherwise A is an *indispensable attribute* in MDT. Let $CORE_\rho(AT)$ is a set of all *indispensable attributes* of AT.

Proposition 1:

$$CORE_\rho(AT) = \bigcap Reduct_\rho(AT)$$

The solution is similar to the solution in[8].

3. Discovery of Decision Rules

In this part, we use the approach that is based on rough set to discover the set of minimal decision rules for MIT. The set consists of the rules, each of which has the smallest number of descriptors such that guarantee to cover the set of all objects in the information system.

3.1 Discernibility Matrix Based Discovery Minimal Decision Rules

Definition 9: Let $MDT = (OB, AT \cup DC, \rho)$ with n objects, m attributes (including decision attributes). $M(T) = [c_{ij}]$ is called discernbility matrix of MDT with elements c_{ij} to be determined as follows:

$c_{ij} = \{v \in A_k | A_k \in AT \cup DC,$
$\qquad v = \{A_k(O_i)/\rho_k - A_k(O_j)/\rho_k\}\}$
$i, j = 1, \ldots, n$
$k = 1, \ldots, m$

c_{ij} is a set of all labels of equivalence classes allowing to distinguish O_i and O_j.

Algorithm 1: To determine minimal decision rules

Input: $MDT = (OB, AT \cup DC, \rho)$
with n objects, m attributes.
Output: Set of minimal decision rules of MDT.
begin
Generate discernibility matrix $M(T) = [c_{ij}]$,
$c_{ij} = \cup\{v | v = \{A_k(O_i)/\rho_k - A_k(O_j)/\rho_k\}\}$
$i, j = 1, .., n; k = 1, .., m$
for every row i of $M(T)$ **do**
for every $v_d = \{v \in D_k, D_k \in DC\}$ **do**
begin
Set $f_d := \bigwedge_j \{ \vee (v \in c_{ij} | v \neq v_d, c_{ij} \supset v_d)\}$
find prime implicants of f_d
Generate decision rules :
"**if** *condition* **then** *decision*",
with *condition* is prime implicant of f_d
and *decision* is v_d
end
Keep only one of the identical rules
end.

Example 2: Let $MDT = (OB, AT \cup DC, \rho)$, in which:

$AT = \{A_1, A_2\}$
$DC = \{A_3\}$
$D_{A_1} = \{0, 1, 2, 3, 4, 5, 6, 7\}$
$D_{A_2} = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$
$D_{A_3} = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$

On D_{A_1}, D_{A_2}, D_{A_3} are equivalence relations $\rho_{A_1}, \rho_{A_2}, \rho_{A_3}$ respectively.

With ρ_{A_1}, we have: $[a_1] = \{0, 1\}$, $[a_2] = \{2, 3\}$, $[a3] = \{4, 5, 6, 7\}$.

With ρ_{A_2}, we have: $[b_1] = \{0, 5, 6, 8\}$, $[b_2] = \{1, 2, 3, 4\}$, $[b_3] = \{7\}$.

With ρ_{A_3}, we have: $[d_1] = \{0, 1, 2\}$, $[d_2] = \{3, 4, 5, 6\}$, $[d_3] = \{7, 8\}$, $[d_4] = \{9, 10\}$.

The figures 2,3,4 show the original MDT, MDT after applying ρ, and the discernbility matrix of the MDT respectively.

Applying the above algorithm, we have:
Consider row 1:

- For $[d1] : f_d := ([a_2] \vee [b_1]) \wedge ([a_1] \vee [b_2])$. Because f_d has four implicants, we have four rules:
  - $[a_1] \wedge [a_2] \Rightarrow [d_1]$
  - $[a_1] \wedge [b_1] \Rightarrow [d_1]$
  - $[a_2] \wedge [b_2] \Rightarrow [d_1]$
  - $[b_1] \wedge [b_2] \Rightarrow [d_1]$
- For $[d_2] : f_d := [a_1]$. Then we have the rule:
  - $[a_1] \Rightarrow [d_2]$

Similarly, we obtain the set of all rules of the MDT above.

Assume that x_5 with condition attribute values $A_1(x_5) = \{0, 1\} \subseteq [a_1]$, $A_2(x_5) = \{0, 5, 6, 8\} \subseteq [b_1]$. Clearly, x_5 matches the rule $[a_1] \wedge [b_1] \Rightarrow [d_1]$. However, in the classical approach no decision is generated with condition attribute values of x_5.

The above algorithm is well applied in $MITs$ that are consistent. However, the algorithm produces a set of rules of which some rules conflict with others if MIT is inconsistent. In order to solve this problem, we propose an algorithm that combines Variable Precision Rough Set model[11] with Apriori algorithm.

3.2 Combining Rough Set Approach With Apriori Algorithm To Discover Minimal Decision Rules

Consider inconsistent MIT. According to the classical rough set approach, decision rules are generated from approximations of decision classes. The set of rules are classified into two basic kinds: certain and approximate. For each decision class, certain decision rules are generated from objects belonging to its lower approximation, approximate decision rules are generated either from upper approximation or from the boundaries of this decision class with other classes[10].

In this combined algorithm, we apply Variable Precision Rough Set model to generate the set of decision rules R with confidence α by calculating positive region of MIT. Then we use Apriori algorithm to produce decision rules AR that satisfy α ; simultaneously remove redundant attribute values of R. After that, we obtain the set of minimal decision rules RULES with the corresponding confidence.

Algorithm 2: To determine minimal decision rules with confidence α

Input: $MDT = (OB, AT \cup DC, \rho)$
    with n objects, m attributes;
    α is confidence.
Output: The set of minimal decision rules
    $RULES$ of MDT with confidence α.
begin
 for every $X \in OB/\rho$ **do**
  begin
   Set $\underline{AT}_\rho X(\alpha) = \bigcup\{Y \in OB/\rho | Y^\alpha \subseteq X\}$
    where $Y^\alpha \subseteq X$ denoting $\frac{|X \cap Y|}{|X|} \leq \alpha$
   Generating decision rule r with
    confidence α from objects of $\underline{AT}_\rho X(\alpha)$
   $R = R \cup r$
  end
 for every $A_i \in AT \cup DC$ **do**
  begin
   //Consider D_i/ρ_i as items
   Apply Apriori in A_i to obtain associate
    rules AR_i with confidence 1
   $AR1 = \cup AR_i$
  end
 //Consider $A_i(O_i)/\rho_i$ as items
 Apply Apriori in MDT to obtain associate
  rules $AR2$ with confidence α
 // The rules in form *left* ⇒ *right*, where *left*
 // and *right* are sets of attribute value classes
 filter ($AR2$, $AR1$)
 filter (R, $AR1$)

OB	A1	A2	A3
x1	{0,2,3}	{0,1,2,5,7}	{1,3}
x2	{1,4,5,6,7}	{2,3,4,7}	{3,7,9,10}
x3	{3}	{1,4,7}	{4,5,6,7,8}
x4	{2,3,4}	{3,5,6,7}	{1,2}

Fig. 2 Original MDT

OB	A1	A2	A3
x1	{[a1], [a2]}	{[b1], [b2], [b3]}	{[d1], [d2]}
x2	{[a1], [a3]}	{[b2], [b3]}	{[d2], [d3], [d4]}
x3	{[a2]}	{[b1], [b3]}	{[d2], [d3]}
x4	{[a2], [a3]}	{[b1], [b2], [b3]}	{[d1]}

Fig. 3 MDT after applying ρ

	x1	x2	x3	x4
x1		[a2],[b1],[d1]	[a1],[b2],[d1]	[a1],[d2]
x2	[a3],[d3],[d4]		[a1],[a3],[b2],[d4]	[a1],[d2],[d3],[d4]
x3	[d3]	[a2],[b1]		[d2],[d3]
x4	[a3]	[a2],[b1],[d1]	[a3],[b2],[d1]	

Fig. 4 Discernbility matrix of MDT

```
// Associate rules AR2 are then minimal
for every ar2_i ∈ AR2 do
  begin
    if left(ar2_i)∩Dom(DC)/ρ ≠ ∅
    then go to next rule
    else
     if right(ar2_i)∩Dom(DC)/ρ ≠ ∅
     then
       begin
        Remove elements of Dom(AT)/ρ
          possibly in right(ar2_i)
        RULES=RULES ∪ ar2_i
       end
     else
       if ar2_i with confidence 1
       then filter(R, ar2_i)
  end
 RULES=RULES∪R
end.
// Procedure use R2 to remove redundant
// values in R1
procedure filter(R1={r1_i}, R2={r2_i})
 begin
  for every r1_i ∈ R1 do
     for every r2_i ∈ R2 do
      begin
        if left(r2_i) ∪ right(r2_i) ⊆ left(r1_i)
         (left(r2_i) ∪ right(r2_i) ⊆ right(r1_i))
         then mark right(r2_i) in left(r1_i)
          (mark right(r2_i) in right(r1_i))
       end
    Remove the remarked elements in r1_i
 end.
```

4. Conclusions

The paper produced a rough set approach to handle MIT. We give the definition of basic notions applied for MIT, in which the definition of discernibility matrix is developed in a quite different way in comparison with the classical one.

Based on the model designed with this approach, two algorithms given here enable to directly discover minimal decision rules from MIT. These decision rules help to reduce the number of redundant attribute values in the rules and simultaneously express the relation between attribute value classes. These relations cannot be discovered in the classical approach. Moreover, the number of rules discovered is reduced but guarantees to cover the set of all objects in the table, including the case applied in single-valued information system.

References

1) Chang-Ling Hsu, "Multi-valued and Multi-labeled Decision Tree Classifiers for Data Mining", *PhD Thesis in Information Management*, China, 2004.
2) Date, C.J., "An Introduction to Database System", 7^{th} *edition, Addison Wesley,* 1999.
3) Fernandez-Baizan, C., Menasalvas, E., Pena, J., Francisco, J., Millan, S., "Using the Apriori Algorithm to Improve Rough Sets Results", *LNAI 2005*, Springer-Verlag, 2000, 291-295.
4) Ho Cam Ha, "An Approach to Extending the Relational Database Model for Handling Incomplete Information",*PhD Thesis in Computer Science*, Hanoi University of Technology, Vietnam, 2001.
5) Kryszkiewicz, M., Rybinski, H., "Mining in Information Systems with Real Value Multi-valued Attributes", *Proc.of Int. Conf. on Information Systems, Analysis and Synthesis, ISAS'96*, Orlando, USA, 22-26 June, 1996.
6) Medical dataset, *PKDD'99 Discovery Challenge*, 1999.
7) Nguyen, S. H., Skowron, A., Synak, P., Wróblewski, J., "Knowledge Discovery In Databases: Rough Set Approach", *Proc. Of The Seventh International Fuzzy Systems Association World Congress, Vol. II, IFSA97*, Prague, Czech Republic,1997, 204-209.
8) Pawlak, Z., "Rough Sets: Theoretical Aspects Of Reasoning About Data", *Kluwer Academic Publishers*, 1991.
9) Pawlak, Z., "Rough Sets And Their Applications", *Computational Intelligence In Theory And Practice*, Springer-Verlag, 2000, 73-91.
10) Stefanowski, J., "On rough set based approaches to induction of decision rules", *Rough Sets in Data Mining and Knowledge Discovery, Vol 1,* (Polkowski L., Skowron A., Eds.), Physica Verlag, Heidelberg, 1998, 500-529.
11) Ziako, W., "Variable Precision Rough Set Model", *Journal of Computer and System Sciences 46,* 1993, 39-59.