Journal of Network and Innovative Computing
ISSN 2160-2174 Volume 3 (2015) pp. 010-018

Relation between definition of Reduct in Set-valued Decision System

Thi Thu Hien Phung[1], Thanh Dai Tran[2]

[1] University of Economic and Technical Industries
456 Minh Khai, Hai Ba Trung District, Ha Noi, Viet Nam
Thuhiencn1@gmail.com

[2] University of Economic and Technical Industries
456 Minh Khai, Hai Ba Trung District, Ha Noi, Viet Nam

***Abstract*: In term of researching attribute reduction based on tolerance rough set, this paper researches relation between definitions of reduct in set-valued decision system. The results help us deeply understand characteristic of reduct's definition in set-valued decision system, also it is a theory foundation in order to create new and efficiency methods.**

For attribute reduction problems, the most important issue is to minimize the time complexity of attribute reduction algorithms. There have been many methods for attribute reduction, but finding reducts in these methods is almost performed on initial object set. In this paper, we propose a method for selection representative object set from initial object set to solve attribute reduction problem in set-valued decision system.

Keywords: *Tolerance rough set, decision system, attribute reduct, reduct.*

I. Introduction

Pawlak's rough set theory [5] suggests unreliable definition instrument and broadly apply for researching data.

In real information systems, attribute value of object might be a valued set. For example, considering an object in the information system 'Nguyen Van A' at attribute 'Language' is 'English, French, Russian', means that Nguyen Van A know language English, or French, or Russian. The information system like that is called set valued information system. In set-valued information system, Yan Yong Guan and co-worker [1] broaden equivalent relation in traditional rough set into tolerance and create tolerance set model.

According to rough set model, some results of reduct's method in set-valued decision system presented [1, 6, 7, 8, 11]. Each method also gives reduct's definition and creates heuristic mathematics to find the best reduct which uses attribute's quality classification standard (or called level of attribute important).

There are some symbolic definition of reduct in set-valued decision system based on rough set model : reduct based on positive region created by positive region's definition in traditional rough set theory [5], reduct based on generalized decision function which uses expansive decision function [2]. According to discernibility matrix and discernibility function in traditional rough set theory [9], authors in [14] show a definition of reduct in set-valued decision system based on discernibility matrix. In the research [6], the author creates discernibility matrix and discernibility function, after that giving a definition of reduct based on discernibility function. In the research [8], authors build contigency table and discernibility function in set-valued decision system, after that building a definition of reduct which used discernibility function.

Reduct is a result of a reduct's method. Thus, it is important to find the relation between reducts which helps to compare and evalute attribute reduction. However, in set-valued decision system, its research's result is limited.

Some results are: the authors in [6] proved that reduct which uses expansive discernibilty function and generalized decision function is similar. The authors in [8] showed that reduct which used discernibilty function and reduct based on discernibility matrix is similar.

Until now, there are many methods to calculate for singer-valued information system [10], but all of them are based on the first objective set. In set-valued information system, the researches [2, 13] presented solutions data squeeze in order to minimize size of the first objective set and reducing time of attribute reduction algorithms. However, a limitation of data squeeze method is no demonstrating attribution reduction in the first objective set which is similar with reduction set after squeeze.

According to the ideas of minimizing the first set's size [3], in this paper we propose methods of selecting representative object set for attribute reduction in set-valued decision system. We demonstrate reduction in the first object set and representative object set which is similar in set-valued decision system. Because of object set's size is smaller than the first object set, so reducing the time complexity of attribute reduction algorithms in representative object set. Representative object set's size selected is depended on characteristic of each information system in real works.

In this paper, section 2 shows basic definition of rough set model in set-valued information system and reduct's definitions. Section 3 presents the relation between reduct's definitions, separating and evaluating methods of attribute reduction. Section 4 proposes methods of selecting representative object set in set-valued decision system. Section 5 is a conclusion and further research.

II. Basic Definitions

In this section, we outline basic concepts of set-valued information systems [12].

Information systems is a tuple $IS=(U,AT)$, where U is a finite set of objects and AT is a finite set of attribute, i.e. the function of form $a:U\to V_a$, where $a(x)\in V_a$ for each $a\in A$. The set V_a is called the domain of attribute $a\in A$. Such information systems will be called the single-valued information systems.

The set-valued information systems have been proposed as a tool to characterize uncertain information or missing information. The idea is to change the definition of attribute. i.e, $a:U\to 2^{V_a}$. In other words $a(x)\subset V_a$ for any $x\in U$ and $a\in A$.

There are many ways to give a semantic interpretation of the set-valued information systems. The most useful interpretations, which are often discussed in literature [12] are:

Conjunctive interpretation: For $x\in U$, and $a\in A$, *a(x)* is interpreted conjunctively. For example, if a is the attribute "familiarizing program languages", then the attribute value $a(u)=\{C++,Java,Pascal\}$ can be interpreted as u knows three program languages: C++, Java and Pascal.

Disjunctive interpretation: For $x\in U$, and $a\in A$, *a(x)* is interpreted disjunctively. For the example above, the value set $a(x)=\{C++,Java,Pascal\}$ can be interpreted as u knows only one of languages C++, Java or Pascal. The value of *a* numeric attribute "ages" b(x) = [20, 25] can be interpreted as the person u has an age between 20 and 25. Incomplete information systems with some unknown attribute values or partial known attribute values are such type of set-valued information systems.

Hybrid interpretation: Combining of the above two models. Some attributes in the information system are interpreted conjunctively, for example an attribute "familiarizing program languages" and some other ones are interpreted disjunctively such as a numeric attribute "ages".

This part presents some basic definitions of tolerance rough set model in set-valued information system [1] and some definitions of reduct in set-valued decision system.

Information system is an tuple $IS=(U,AT)$, where U is a finite set of objects and AT is finite set of attribute. The value of an attribute $a\in AT$ at an objects $u\in U$ is denoted as *a(u)*. Each sub-set $A\subseteq AT$ determines one equivalence relation:

$$IND(A)=\{(u,v)\in U\times U\,|\,\forall a\in A,\ a(u)=a(v)\}.$$

Partition of U generated by a relation *IND(A)* is denoted as U/A and is denoted as $[u]_A$, while $[u]_A=\{v\in U\,|\,(u,v)\in IND(A)\}$. It is easy to see that $[u]_A=\cap[u]_{\{a\}}$ with all equivalence class in the partition U/A which includes $u\in U$ $a\in A$.

Considering information system $IS=(U,AT)$, if existing $u\in U$ in order to *a(u)* contain *at least two values* then $IS=(U,AT)$ is called *set-valued information system*. Each sub attribute set $A\subseteq AT$, one binary relationship on U is defined as following below:

$$T_A=\{(u,v)\in U\times U\,|\ \forall a\in A, a(u)\cap a(v)\neq\varnothing\}.$$

It is easy to see that T_A is not an equivalent relation because of being reflected, symmetric, no transitive attitude. T_A is called *tolerance relationship* and $T_A=\bigcap_{a\in A}T_a$.

Setting $T_A(u)=\{v\in U\,|\,(u,v)\in T_A\}$, $T_A(u)$ is called a classification of tolerance. The symbol $U/T_A=\{T_A(u)|\ u\in U\}$ presents all classification of tolerance which created by relationship T_A, when U/T_A creates a site of U because of classification in U/T_A can be mix and $\bigcup_{u\in U}T_A(u)=U$. It is easy too see if $B\subseteq A$ then $T_A(u)\subseteq T_B(u)$ with all $u\in U$.

Similar as traditional rough set theory [4], with $B\subseteq A$, $X\subseteq U$, *B-lower approximation* of X is set $\underline{B}X=\{u\in U\,|\,T_B(u)\subseteq X\}=\{u\in X\,|\,T_B(u)\subseteq X\}$.

B-upper approximation of X is set $\overline{B}X=\{u\in U\,|\,T_B(u)\cap X\neq\varnothing\}=\bigcup\{T_B(u)|u\in U\}$, *B- Border region* of *X* is set $BN_P(X)=\overline{P}X-\underline{P}X$.

With all approximate set, calling *B-positive region* for {d} is set: $POS_B(\{d\})=\bigcup_{X\in U/\{d\}}(\underline{B}X)$.

Set-valued decision system is set-valued information system $DS=(U,AT\cup\{d\})$ inside AT is conditional attribution and *d* is decision attribution, with an assumption d(u) contains *one value* with all $u\in U$. Similar as set uncompleted decision system [2], with $u\in U$, $\partial_{AT}(u)=\{d(v)|v\in T_{AT}(u)\}$ is called *expansive decision set* of object *u* on attribut set *AT*.

If $|\partial_{AT}(u)|=1$ with all $u\in U$ then *DS* is *consistent*, opposite *DS* is *inconsistent*. According to positive region theory, *DS* is consistent if and only if $POS_{AT}(\{d\})=U$, opposite *DS* is inconsistent.

Reduct is a basic theory in rough set theory. In general, reduct is a smallest sub-set of condition attribute which keep the classific information of decision set. Next, the authors present some definition about reduct of set-valued decision system which is used in this article.

By using positive region theory in set-valued decision system, the authors give a definition of reduct based on positive region theory.

Definition 2.1. For set-valued decision system $DS=(U, AT\cup\{d\})$. If $R\subseteq AT$ satisfies:

(1) $POS_R(\{d\})=POS_{AT}(\{d\})$

(2) $\forall R^{'}\subset R$, $POS_{R^{'}}(\{d\})\neq POS_{AT}(\{d\})$.

then R is called as a *reduct of DS based on positive region*.

Using expansive decision function in [2], reduct which based on set valued decision is defined as:

Definition 2.2. For set-valued decision system $DS=(U, AT\cup\{d\})$. If $R\subseteq AT$ satisfies:

(1) $\partial_R(u)=\partial_{AT}(u)$ with all $u\in U$.

(2) $\forall R^{'}\subset R$, existing $u\in U$ in order to $\partial_{R^{'}}(u)\neq\partial_{AT}(u)$

then R is called a reduct of *DS based on expansive decision system*.

According to rough set model in set-valued information system, the authors in [14] give a definition of reduct in discernibility matrix. Discernibility matrix of set-valued decision system DS is $M_{DS}=\left[m_{i\,j}\right]_{n\times n}$, elements $m_{i\,j}$ determined as:

$$m_{i\,j}=\begin{cases}\left\{a\,\middle|\; a\in AT,\;\; (u_i,u_j)\notin T_{AT}\right\} & d(u_i)\neq d(u_j)\\ \varnothing & otherwise\end{cases}$$

Definition 2.6. For set-valued decision system $DS=(U, AT\cup\{d\})$ and discernibility matrix $M_{DS}=\left[m_{i\,j}\right]_{n\times n}$. If $R\subseteq C$ satisfies:

(1) $R\cap m_{i\,j}\neq\varnothing$ with all $m_{i\,j}\neq\varnothing$

(2) With all $r\in R$, $R\setminus\{r\}$ do not satisfy (1)

then R is called as a *reduct of DS based on* discernibility matrix.

Approaching to matrix way, the authors in [8] build discernibility function from contingency table and gave a definition of *reduct based on discernibility*. Accroding to definition of discernibility matrix and discernibility function in traditional rough set theory [9], the author [6] gave a definition of *reduct based on generalized discernibility function.*

III. Relation between reduct' definitions

In this part, the authors evaluate and research the relation between reduct' definition of set-valued decision system in part 2. In order to briefing, the authors used reduct symbols as following:

Table 3.1. *Symbol of reduct in set-valued decision system*

Reduct symbol	Brief
R_P	*Reduct based on positive region*
R_∂	*Reduct based on expansive decision function*
R_M	*Reduct based on discernibility matrix*
R_{DF}	*Reduct based on generalized discernibility function*
R_{CF}	*Reduct based on discernibility function*

In set-valued decision system, the author [6] proved R_{DF} is similar with R_∂, the authors [8] showed R_{CF} is similar with R_M. Nextly, the authors research a rest of relation between reducts in terms of *consistent* or *inconsistent* set valued decision system.

3.1. Relation between R_∂ and R_P

Clause 3.1. For set-valued decision system $DS=(U, AT\cup\{d\})$ and $R\subseteq AT$. If $\partial_R(u)=\partial_{AT}(u)$ with all $u\in U$ then $POS_R(\{d\})=POS_{AT}(\{d\})$.

Proof: Assuming $POS_R(\{d\})\neq POS_{AT}(\{d\})$, when surely existing $u_0\in U$ in order to $u_0\in POS_{AT}(\{d\})$ and $u_0\notin POS_R(\{d\})$. From $u_0\in POS_{AT}(\{d\})$ infering $|\partial_{AT}(u_0)|=1$, from $u_0\notin POS_R(\{d\})$ infering $|\partial_R(u_0)|\neq 1$. Thus, $|\partial_R(u_0)|\neq|\partial_{AT}(u_0)|$.

Because of $T_{AT}(u_0)\subseteq T_R(u_0)$ that $\partial_{AT}(u_0)\subseteq\partial_R(u_0)$, combining with $|\partial_R(u_0)|\neq|\partial_{AT}(u_0)|$ infering $\partial_{AT}(u_0)\neq\partial_R(u_0)$. This is contractive with a condition $\partial_R(u)=\partial_{AT}(u)$ with all $u\in U$. So, assuming it is wrong and conclusing if $\partial_R(u)=\partial_{AT}(u)$ with all $u\in U$ then $POS_R(\{d\})=POS_{AT}(\{d\})$.

Note: If *DS* is inconsistent then another way of clause 3.1 is not satisfied. It is illustrated by example 3.1.

Example 3.1. Considering set-valued decision system in table 3.2.

Table 3.2. *Set- valued decision system in example 3.1*

U	a_1	a_2	a_3	d
u_1	1	1	0	0
u_2	1	1	0	1
u_3	1	{1, 2}	0	0
u_4	{1, 2}	2	2	2
u_5	2	2	{0, 2}	0

Getting $\partial_{AT}(u_1)=\partial_{AT}(u_2)=\partial_{AT}(u_3)=\{0,1\}$, $\partial_{AT}(u_4)=\partial_{AT}(u_5)=\{0,2\}$. So *DS* is inconsistent and it is easy to see $POS_{AT}(\{d\})=\varnothing$. Considering $R=\{a_1,a_2\}$, it is easy to see $POS_R(\{d\})=POS_{AT}(\{d\})=\varnothing$.

However, $\partial_{AT}(u_3)=\{0,1\}$ and $\partial_R(u_3)=\{0,1,2\}$, thus $\partial_R(u_3)\neq\partial_{AT}(u_3)$.

Clause 3.1 shows that *if R_∂ is a reduct based on expansive decision system then existing $R_P \subseteq R_\partial$ with R_P is a reduct in positive region*.

If DS is consistent then $POS_R(\{d\})=POS_{AT}(\{d\})=U$, means that with all $u\in U$ having $|\partial_R(u)|=|\partial_{AT}(u)|=1$, or $\partial_R(u)=\partial_{AT}(u)$. So, $\partial_R(u)=\partial_{AT}(u)$ with all $u\in U$ if and only if $POS_R(\{d\})=POS_{AT}(\{d\})$, this is mean that *$R_\partial$ is equivalent with R_P*.

3.2. Relation between R_M and R_∂

In order to research relation between R_M and R_∂, firstly the authors demonstrated lemma as below:

Lemma 3.1. For set-valued decision system $DS=(U,AT\cup\{d\})$, $R\subseteq AT$ and $M_{DS}=\left[m_{ij}\right]_{n\times n}$ is discernibility matrix of DS. When this is a condition $R\cap m_{ij}\neq\varnothing$ with $\forall m_{ij}\neq\varnothing$ if and only if $T_R(u_i)-T_{R\cup\{d\}}(u_i)=T_{AT}(u_i)-T_{AT\cup\{d\}}(u_i)$ with $\forall u_i\in U$.

Proof: Considering set-valued decision system $DS=(U,AT\cup\{d\})$ with $U=\{u_1,u_2,...,u_n\}$ and $R\subseteq AT$.

1) Demonstrating if $T_R(u_i)-T_{R\cup\{d\}}(u_i)=T_{AT}(u_i)-T_{AT\cup\{d\}}(u_i)$ with $\forall u_i\in U$ then $R\cap m_{ij}\neq\varnothing$ with all $m_{ij}\neq\varnothing$.

Assuming to exist $m_{i_0 j_0}\neq\varnothing$ in order to $R\cap m_{i_0 j_0}=\varnothing$. When existing u_{i_0}, $u_{j_0}\in U$ in order to $d(u_{i_0})\neq d(u_{j_0})$. u_{i_0}, u_{j_0} do not distinguish with each other because of attributes in R and u_{i_0}, u_{j_0} distinguish with each other because of attribute in $AT-R$, means that $u_{j_0}\notin T_{AT}(u_{i_0})$ and $u_{j_0}\in T_R(u_{i_0})$. From $u_{j_0}\notin T_{AT}(u_{i_0})$ infering $u_{j_0}\notin T_{AT}(u_{i_0})-T_{AT\cup\{d\}}(u_{i_0})$ (*).

From $u_{j_0}\in T_R(u_{i_0})$ và $d(u_{i_0})\neq d(u_{j_0})$ infering $u_{j_0}\in T_R(u_{i_0})-T_R(u_{i_0})\cap T_{\{d\}}(u_{i_0})$ or $u_{j_0}\in T_R(u_{i_0})-T_{R\cup\{d\}}(u_{i_0})$.

(**). From (*) and (**) infering $T_R(u_i)-T_{R\cup\{d\}}(u_i)\neq T_{AT}(u_i)-T_{AT\cup\{d\}}(u_i)$, this is contractive with the assumption. So that, assumption is wrong and getting a demonstration.

2) Otherwise, the authors need to demonstrate if $R\cap m_{ij}\neq\varnothing$ with all $m_{ij}\neq\varnothing$ then $T_R(u_i)-T_{R\cup\{d\}}(u_i)=T_{AT}(u_i)-T_{AT\cup\{d\}}(u_i)$ với $\forall u_i\in U$.

Assuming to exist u_{i_0} in order to $T_R(u_{i_0})-T_{R\cup\{d\}}(u_{i_0})\neq T_{AT}(u_{i_0})-T_{AT\cup\{d\}}(u_{i_0})$. Because of $T_{AT}(u_{i_0})-T_{AT\cup\{d\}}(u_{i_0})\subseteq T_R(u_{i_0})-T_{R\cup\{d\}}(u_{i_0})$ so existing $u_{j_0}\in U$ in order to $u_{j_0}\in T_R(u_{i_0})-T_{R\cup\{d\}}(u_{i_0})$ and $u_{j_0}\notin T_{AT}(u_{i_0})-T_{AT\cup\{d\}}(u_{i_0})$.

From $u_{j_0}\in T_R(u_{i_0})-T_{R\cup\{d\}}(u_{i_0})$ infering $d(u_{j_0})\neq d(u_{i_0})$, combining with $u_{j_0}\notin T_{AT}(u_{i_0})-T_{AT\cup\{d\}}(u_{i_0})$ infering $u_{j_0}\notin T_{AT}(u_{i_0})$. According to discernibility matrix theory, existing $m_{i_0 j_0}\neq\varnothing$ in order to with all $a\in m_{i_0 j_0}$ then $a\notin R$ (because of $u_{j_0}\in T_R(u_{i_0})$)), means that $R\cap m_{i_0 j_0}=\varnothing$. This is contractive with the condition $R\cap m_{ij}\neq\varnothing$ with all $m_{ij}\neq\varnothing$. Thus, an assumption is false and getting a demonstration.

From 1) and 2) conclused that $R\cap m_{ij}\neq\varnothing$ with $\forall m_{ij}\neq\varnothing$ *if and only if* $T_R(u_i)-T_{R\cup\{d\}}(u_i)=T_{AT}(u_i)-T_{AT\cup\{d\}}(u_i)$ with $\forall u_i\in U$.

Clause 3.2. *For set-valued decision system $DS=(U,AT\cup\{d\})$ and $R\subseteq AT$. If $R\cap m_{ij}\neq\varnothing$ with $\forall m_{ij}\neq\varnothing$ then $\forall u\in U$, $\partial_R(u)=\partial_A(u)$.*

Proof: Considering set-valued decision system $DS=(U,AT\cup\{d\})$ with $U=\{u_1,u_2,...,u_n\}$ and $R\subseteq AT$. According to lemma 3.1, condition $R\cap m_{ij}\neq\varnothing$ with $\forall m_{ij}\neq\varnothing$ is equivalent with:

$T_R(u_i)-T_{R\cup\{d\}}(u_i)=T_{AT}(u_i)-T_{AT\cup\{d\}}(u_i)$ with $\forall u_i\in U$

On the other hand,

$$T_R(u_i)=\left(T_R(u_i)\cap T_{\{d\}}(u_i)\right)\cup\left(T_R(u_i)-\left(T_R(u_i)\cap T_{\{d\}}(u_i)\right)\right)$$

$$T_{AT}(u_i)=\left(T_{AT}(u_i)\cap T_{\{d\}}(u_i)\right)\cup\left(T_{AT}(u_i)-\left(T_{AT}(u_i)\cap T_{\{d\}}(u_i)\right)\right)$$

$$d_i=d(u_i),\ R_i=\left\{d(v_i)\middle|v_i\in T_R(u_i)-\left(T_R(u_i)\cap T_{\{d\}}(u_i)\right)\right\},$$

$$AT_i=\left\{d(v_i)\middle|v_i\in T_{AT}(u_i)-\left(T_{AT}(u_i)\cap T_{\{d\}}(u_i)\right)\right\}$$

Getting:

$$\partial_R(u_i)=\left\{\begin{array}{l} d(v_i)\middle|v_i\in\left(T_R(u_i)\cap T_{\{d\}}(u_i)\right)\cup \\ \left(T_R(u_i)-\left(T_R(u_i)\cap T_{\{d\}}(u_i)\right)\right)\end{array}\right\}=\{d_i\}\cup R_i$$

$$\partial_{AT}(u_i)=\left\{\begin{array}{l} d(v_i)\middle|v_i\in\left(T_{AT}(u_i)\cap T_{\{d\}}(u_i)\right)\cup \\ \left(T_{AT}(u_i)-\left(T_{AT}(u_i)\cap T_{\{d\}}(u_i)\right)\right)\end{array}\right\}=\{d_i\}\cup AT_i$$

From formula (1) infering $R_i = AT_i$, so $\partial_R(u_i) = \partial_{AT}(u_i)$ with all $u_i \in U$.

Note: If *DS* is inconsistent then oposite site of clause 3.2 is not satisfied because of the condition $\forall u_i \in U, \partial_R(u_i) = \partial_{AT}(u_i)$ only keep a general decision formular $\partial_R(u_i)$ of tolerance classification $T_R(u_i)$, condition $R \cap m_{ij} \neq \varnothing$ with $\forall m_{ij} \neq \varnothing$ (or $T_R(u_i) - T_{R\cup\{d\}}(u_i) = T_{AT}(u_i) - T_{AT\cup\{d\}}(u_i)$ with $\forall u_i \in U$) keep inconsistent objects with u_i of tolerance classification $T_R(u_i)$ (strict condition). That is illustrated by example 3.2 below.

Example 3.2. Considering set-valued decision system in table 3.3.

Table 3.3. *Set valued decision system in example 3.2*

U	a_1	a_2	a_3	d
u_1	1	{1, 2}	0	1
u_2	1	1	0	0
u_3	1	{1, 2}	0	0
u_4	{1, 2}	2	2	0
u_5	2	2	{0, 2}	1

Getting: $T_{AT}(u_1) = T_{AT}(u_2) = T_{AT}(u_3) = \{u_1, u_2, u_3\}$, $T_{AT}(u_4) = T_{AT}(u_5) = \{u_4, u_5\}$,

$\partial_{AT}(u_1) = \partial_{AT}(u_2) = \partial_{AT}(u_3) = \partial_{AT}(u_4) = \partial_{AT}(u_5) = \{0,1\}$

So, *DS* is inconsistent. Considering $R = \{a_1, a_2\}$ got $T_R(u_1) = T_R(u_3) = \{u_1, u_2, u_3, u_4\}$, $T_R(u_2) = \{u_1, u_2, u_3\}$, $T_R(u_4) = \{u_1, u_3, u_4, u_5\}$, $T_R(u_5) = \{u_4, u_5\}$,
$\partial_R(u_1) = \partial_R(u_2) = \partial_R(u_3) = \partial_R(u_4) = \partial_R(u_5) = \{0,1\}$

Thus, with $\forall u_i \in U, i = 1..5$, $\partial_R(u_i) = \partial_{AT}(u_i)$. On the other hand, indiscernibility matrix of *DS* is:

$$M_{DS} = \begin{bmatrix} 0 & 0 & 0 & \{a_3\} & 0 \\ 0 & 0 & 0 & 0 & \{a_1, a_2\} \\ 0 & 0 & 0 & 0 & \{a_1\} \\ \{a_3\} & 0 & 0 & 0 & 0 \\ 0 & \{a_1, a_2\} & \{a_1\} & 0 & 0 \end{bmatrix}$$

Clearly, $\{a_1, a_2\} \cap \{a_3\} = \varnothing$.

According to clause 3.2 *if R_M is a reduct based on discernibility matrix then existing $R_\partial \subseteq R_M$ with R_∂ is a reduct based on expansive decision function*.

If *DS* is consistent, from condition $\forall u_i \in U, |\partial_R(u_i)| = |\partial_{AT}(u_i)| = 1$ infering $T_R(u_i) = T_{R\cup\{d\}}(u_i)$ and $T_A(u_i) = T_{A\cup\{d\}}(u_i)$ with $\forall u_i \in U$, so $T_R(u_i) - T_{R\cup\{d\}}(u_i) = T_{AT}(u_i) - T_{AT\cup\{d\}}(u_i)$ with $\forall u_i \in U$.

According to lemma 3.1, getting $R \cap m_{ij} \neq \varnothing$ with $\forall m_{ij} \neq \varnothing$. There for**,** $R \cap m_{ij} \neq \varnothing$ với $\forall m_{ij} \neq \varnothing$ if and only if $\forall u_i \in U, \partial_R(u_i) = \partial_{AT}(u_i)$, means is R_M *is equivalent with* R_∂.

3.3 Summarizing the relation between reduct' definitions of set-valued decision system

Based on the researches, the authors summarize the relation between reduct' definitions of set-valued decision system:

If decision system is consistent, reducts R_P, R_∂, R_M, R_{DF}, R_{CF} are similar.

If decision table is inconsistent, the relation between presented by a model as followed:

Model 3.1*.* *Relation between reducts of set-valued decision system*

$$\boxed{R_P} \xrightarrow[\subseteq]{} \boxed{R_\partial = R_{DF}} \xrightarrow[\subseteq]{} \boxed{R_\partial = R_{DF}}$$

Group 1: Including reduct R_P.

G*roup 2*: Including reducts R_∂, R_{DF}.

Group 3: Including reducts R_M, R_{CF}.

Relation of reducts into groups as:

If R_3 is a reduct of group 3 then existing a reduct R_2 belong to group 2 and a reduct R_1 belong to group 1 in order to $R_1 \subseteq R_2 \subseteq R_3$.

3.4 Evaluating methods of attribute reduction in set-valued decision system based on reduct

After giving reduct's definition, methods of attribute reduction build heuristic algorithm to find the best reduct based on level of important standard of attribute, or quality of attribute classification.

With set-valued consistent decision system, the best reducts of three methods are the same, so they have similar quality of classifications. With set-valued inconsistent decision system, the authors evaluate three methods based on reduct classification's quality standard.

Assumming R_{MBest} is a reduct of method which belong to Group 3 (found by heuristic algorithms which used discernibility matrix, or *discernibility function*). Following the resuts, existing a reduct based on expansive decision function R_∂ in order to $R_\partial \subseteq R_{MBest}$ (R_∂ is minimum than R_{MBest}).

Assumming is the best reduct of the second method

($R_{\partial Best}$ found by heuristic algorithms which used *expansive decision function* or *expansive discernibility function*). There are two cases.

- If $R_{\partial Best}$ is R_{∂} ($R_{\partial Best} = R_{\partial}$) then $R_{\partial Best} \subseteq R_{MBest}$, means that $R_{\partial Best}$ is minimum than R_{MBest}. Mean that , or quality classification of $R_{\partial Best}$ is better than R_{MBest}.

- If $R_{\partial Best}$ is different with R_{∂} then quality classification of $R_{\partial Best}$ is better than R_{∂} caused of quality classification of $R_{\partial Best}$ is the best one. On the other hand, caused of $R_{\partial} \subseteq R_{MBest}$ so R_{∂} is better than R_{MBest} about quality classification. Thus, $R_{\partial Best}$ is better than R_{MBest} about quality classification.

Therefore, in both two cases, quality classification of $R_{\partial Best}$ is better than R_{MBest}. Thus, to summerize, the methods of *group 2* is more efficiency than method of *group 3* which followed by quality classification of reduct standard.

Similarity, the methods of *group 1* is more efficiency than method of *group 2* which followed by quality classification of reduct standard. Thus, the method of finding reduct based on *generalized discernibility fuction* in [6] is better than *discernibility function* in [8] which followed by quality classification of reduct standard.

IV. Selecting of representative object set for attribute reduction in set-valued decision system

Selecting a representative object set is essentially data preprocessing step in the data mining and machine learning. Instead of finding a reduct on overall initial object set, we find the reduct based on representative object set and prove theorically that the reduct obtained from the representative object equals to the reduct obtained from the initial representative object set. Because the size of the representative object is much smaller that of initial object set, the time complexity of attribute reduction algorithm is much smaller. Representative object set consists of representative objects; each object is selected as follow:

Let us consider the set-valued decision system $DS = (U, AT \cup \{d\})$, based on attribute set *AT*, we first partition the initial set of objects *U* into equivalence classes. Two objects $u, v \in U$ belong to the same equivalence class if $T_{\{a\}}(u) = T_{\{a\}}(v)$ for all $a \in AT$. We calculate the partition $X_i / \{d\} = \{Y_1, ..., Y_l\}$. With equivalence class $Y_j \in X_i / \{d\}$, we select a representative object, without generality, we select the first object which is a representative object. Then, selected object set is representative object set.

The algorithm for selecting representative object set in set-valued decision system is presented as follow:

Algorithm 1. Selecting representative object set of a set-valued decision system.

Input: The initial set-valued decision system $DS = (U, AT \cup \{d\})$ with $U = \{u_1, ..., u_n\}$, $AT = \{a_1, ..., a_m\}$.

Output: The representative set-valued decision system $DS_P = (U_P, AT \cup \{d\})$ with $U_P \subseteq U$ is representative object set.

Step 1: Set $U_P = \varnothing$;

Step 2: For any $a_i \in AT, i = 1..m$, calculate partition $U / \{a_i\} = \{[u]_{\{a_i\}} \mid u \in U\}$ where

$$[u]_{\{a_i\}} = \{v \in U \mid T_{\{a_i\}}(u) = T_{\{a_i\}}(v)\}.$$

Step 3: Calculate partition $U / AT = \{[u]_{AT} \mid u \in U\}$ where $[u]_{AT} = [u]_{\{a_1\}} \cap ... \cap [u]_{\{a_m\}} = \bigcap_{i=1}^{m} [u]_{\{a_i\}}$.

Assume that $U / AT = \{X_1, ..., X_k\}$;

Step 4: For $X_i \in U / AT, i = 1..k$, do Step 4.1 and 4.2 as follow:

Step 4.1. Calculate partition $X_i / \{d\} = \{[u]_{\{d\}} \mid u \in X_i\}$

with $[u]_{\{d\}} = \{v \in X_i \mid d(u) = d(v)\}$.

Assume that $X_i / \{d\} = \{Y_1, ..., Y_l\}$

and $Y_j = \{u_{j_1}, ..., u_{j_o}\}$ with $j = 1..l$.

Step 4.2. With each $Y_j \in X_i / \{d\}$, $j = 1..l$, set $U_P := U_P \cup \{u_{j_1}\}$;

Step 5: Return $DS_P = (U_P, AT \cup \{d\})$;

Assume that *k* is conditional attribute and *n* is object. At step 2, for each $a_i \in A, i = 1..m$, the time complexity to calculate $T_{\{a_i\}}(u), u \in U$ is $O(n^2)$, the time complexity to calculate $U / \{a_i\}$ is $O(n \log n)$. Thus, the complexity of Step 2 is $O(kn^2)$. The complexity of Step 3 is $O(n)$. The complexity of Step 4 is $O(n \log n)$.

Therefore, the time complexity of algorithm 1 is $O(kn^2)$.

Example 4.1. Considering set-valued decision system $DS = (U, AT \cup \{d\})$ in Table 4.1

Table 4.1. *Set-valued decision system*

U	a_1	a_2	a_3	a_4	d
u_1	{1}	{ 1}	{1}	{0}	1
u_2	{0}	{0, 1}	{1}	{0}	1
u_3	{0, 1}	{0, 1}	{0}	{1}	0
u_4	{1}	{0, 1}	{1}	{1}	1
u_5	{0, 1}	{0, 1}	{1}	{1}	2
u_6	{0}	{1}	{1}	{0, 1}	1
u_7	{0, 1}	{1}	{0}	{0, 1}	0
u_8	{0}	{1}	{1}	{0}	1
u_9	{0, 1}	{0, 1}	{0}	{1}	0

We have:

$U/AT=\{\{u_1\},\{u_2,u_8\},\{u_3,u_9\},\{u_4\},\{u_5\},\{u_6\},\{u_7\}\}$

$T_{\{a_1\}}(u_1)=T_{\{a_1\}}(u_4)=\{u_1,u_3,u_4,u_5,u_7,u_9\}$,

$T_{\{a_1\}}(u_3)=T_{\{a_1\}}(u_5)=T_{\{a_1\}}(u_7)=T_{\{a_1\}}(u_9)=U$,

$T_{\{a_1\}}(u_2)=T_{\{a_1\}}(u_6)=T_{\{a_1\}}(u_8)=\{u_2,u_3,u_5,u_6,u_7,u_8,u_9\}$.

Consequently:

$U/\{a_1\}=\{\{u_1,u_4\},\{u_2,u_6,u_8\},\{u_3,u_5,u_7,u_9\}\}$.

Similarly, we have $U/\{a_2\}=U$,

$U/\{a_3\}=\{\{u_1,u_2,u_4,u_5,u_6,u_8\},\{u_3,u_7,u_9\}\}$,

$U/\{a_4\}=\{\{u_1,u_2,u_8\},\{u_3,u_4,u_5,u_9\},\{u_6,u_7\}\}$

Then we have

$$U/AT=\{\{u_1\},\{u_2,u_8\},\{u_3,u_9\},\{u_4\},\{u_5\},\{u_6\},\{u_7\}\}$$
$$=\{X_1,X_2,X_3,X_4,X_5,X_6,X_7\}$$

Calculate $X_1/\{d\}=\{\{u_1\}\}$, so u_1 is selected and $U_P:=\{u_1\}$.

Calculate $X_2/\{d\}=\{\{u_2,u_8\}\}$, so u_2 is seclected and $U_P:=\{u_1,u_2\}$.

Calculate $X_3/\{d\}=\{\{u_3,u_9\}\}$, so u_3 is selected and $U_P:=\{u_1,u_2,u_3\}$.

Calculate $X_4/\{d\}=\{\{u_4\}\}$, so u_4 is selected and $U_P:=\{u_1,u_2,u_3,u_4\}$.

Calculate $X_5/\{d\}=\{\{u_5\}\}$, so u_5 is selected and $U_P:=\{u_1,u_2,u_3,u_4,u_5\}$.

Calculate $X_6/\{d\}=\{\{u_6\}\}$, so u_6 is selected and $U_P:=\{u_1,u_2,u_3,u_4,u_5,u_6\}$.

Calculate $X_7/\{d\}=\{\{u_7\}\}$, so u_7 is selected and $U_P:=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7\}$.

Thus, representative object set is selected $U_P:=\{u_1,u_2,u_3,u_4,u_5,u_6,u_7\}$ and set-valued decision system $DS_P=(U_P,AT\cup\{d\})$ is selected in table 4.2.

Table 4.2. *Represetative set-valued decision system from table 4.1*

U	a_1	a_2	a_3	a_4	d
u_1	{1}	{ 1}	{1}	{0}	1
u_2	{0}	{0, 1}	{1}	{0}	1
u_3	{0, 1}	{0, 1}	{0}	{1}	0
u_4	{1}	{0, 1}	{1}	{1}	1
u_5	{0, 1}	{0, 1}	{1}	{1}	2
u_6	{0}	{1}	{1}	{0, 1}	1
u_7	{0, 1}	{1}	{0}	{0, 1}	0

In the first set-valued decision table $DS=(U,AT\cup\{d\})$ and representative set-valued decision table $DS_P=(U_P,AT\cup\{d\})$, we prove the following lemma:

Lemma 2. If $u_p\in U$ is a representative object in $DS=(U,AT\cup\{d\})$ such that $\partial_B(u_p)\neq\partial_{AT}(u_p)$ where $B\subset AT$, then $\partial_B(u_p)\neq\partial_{AT}(u_p)$ in $DS_P=(U_P,AT\cup\{d\})$ where $u_p\in U_p$.

***Proof*:** On $DS=(U,AT\cup\{d\})$, from assumption $\partial_B(u_p)\neq\partial_{AT}(u_p)$ we have $T_B(u_p)\neq T_{AT}(u_p)$. Suppose that $T_B(u_p)=T_{AT}(u_p)\cup Y$, then there exists $y\in Y$ such that $d(y)\notin\partial_{AT}(u_p)$, we have $y\notin[u_p]_{AT}$.

1) If y is a representative object, it means that $y=y_p$, then on $DS_P=(U_P,AT\cup\{d\})$, $d(y_p)\notin\partial_{AT}(u_p)$ and $d(y_p)\in\partial_B(u_p)$, so we can conclude that $\partial_B(u_p)\neq\partial_{AT}(u_p)$.

2) If y is not a representative object, suppose that y_p is a representative object of a equivalence class which contains y, based on the method of building representative object set, we have $d(y_p)=d(y)$, from $d(y)\notin\partial_{AT}(u_p)$, on $DS_P=(U_P,AT\cup\{d\})$, we have $d(y_p)\notin\partial_{AT}(u_p)$ **(i)**.

In addition, on $DS=\left(U,AT\cup\{d\}\right)$ we have $d\left(y_p\right)=d\left(y\right)$, from $d\left(y\right)\in\partial_B\left(u_p\right)$ we have $d\left(y_p\right)\in\partial_B\left(u_p\right)$.

Thus, on $DS_P=\left(U_P,AT\cup\{d\}\right)$ we have $d\left(y_p\right)\in\partial_B\left(u_p\right)$ **(ii)**.

From **(i)** and **(ii)** we conclude that $\partial_B\left(u_p\right)\neq\partial_{AT}\left(u_p\right)$.

Thus, both two situation 1) and 2) we have $\partial_B\left(u_p\right)\neq\partial_{AT}\left(u_p\right)$ on $DS_P=\left(U_P,AT\cup\{d\}\right)$.

Next, we prove that the reduct of initial set-valued decision system and the reduct of representative set-valued decision system are the same.

Assumption $R\subseteq AT$ is a reduct of the initial set-valued decision system $DS=\left(U,AT\cup\{d\}\right)$, then $\partial_R(u)=\partial_{AT}(u)$ for any $u\in U$ and $\forall B\subset R$ there exists $u\in U$ such that $\partial_R(u)\neq\partial_{AT}(u)$.

1) From $\partial_R(u)=\partial_{AT}(u)$ for any $u\in U$ on $DS=\left(U,AT\cup\{d\}\right)$ we have $\partial_R(u_p)=\partial_{AT}(u_p)$ for any $u_p\in U_P$ on $DS_P=\left(U_P,AT\cup\{d\}\right)$.

2) Without generality, assume that $B\subset R$ and exist $u\in U$ on $DS=\left(U,AT\cup\{d\}\right)$ such that $\partial_B(u)\neq\partial_{AT}(u)$.

If u is a representative object then $u=u_p$ and $\partial_B\left(u_p\right)\neq\partial_{AT}\left(u_p\right)$ on $DS=\left(U,AT\cup\{d\}\right)$, according to lemma 2 we have $\partial_B\left(u_p\right)\neq\partial_{AT}\left(u_p\right)$ on $DS_P=\left(U_P,AT\cup\{d\}\right)$ **(i)**.

If u is not a representative object then on $DS=\left(U,AT\cup\{d\}\right)$, assume that u_p is a representative object of equivalence class $\left[u_p\right]_{AT}$ wich contains u and u_p, then $\left[u_p\right]_{AT}=\left[u\right]_{AT}$. Because of $B\subset R\subseteq AT$, so from $\left[u_p\right]_{AT}=\left[u\right]_{AT}$ we have $\left[u_p\right]_B=\left[u\right]_B$. From $\left[u_p\right]_{AT}=\left[u\right]_{AT}$ we have $T_{AT}\left(u_p\right)=T_{AT}\left(u\right)$, thus $\partial_{AT}\left(u_p\right)=\partial_{AT}\left(u\right)$. From $\left[u_p\right]_B=\left[u\right]_B$, by similar method we have $\partial_B\left(u_p\right)=\partial_B\left(u\right)$. According to the assumption, $\partial_B\left(u\right)\neq\partial_{AT}\left(u\right)$ so we have $\partial_B\left(u_p\right)\neq\partial_{AT}\left(u_p\right)$ on $DS=\left(U,AT\cup\{d\}\right)$, according to lemma 2, we have $\partial_B\left(u_p\right)\neq\partial_{AT}\left(u_p\right)$ on $DS_P=\left(U_P,AT\cup\{d\}\right)$. **(ii)**

Thus, both two situations **(i)** and **(ii)** we have $\partial_B\left(u_p\right)\neq\partial_{AT}\left(u_p\right)$ on $DS_P=\left(U_P,AT\cup\{d\}\right)$, so we can conclude that there exists $B\subset R$ such that $\partial_B\left(u_p\right)\neq\partial_{AT}\left(u_p\right)$.

From 1) and 2) according to the definition, we have $R\subseteq AT$ is a reduct of representative set-valued information system $DS_P=\left(U_P,AT\cup\{d\}\right)$.

V. Conclusion

In this paper, we presented a method of selecting of representative object set for attribute reduction in set-valued decision system. The authors evaluated and researched relation between reduct' definitions in set-valued decision system. The authors relised that reduct is similar in term of consistent decision table. For inconsistent decision table, the authors separated reducts into *three group* and presented the relation of reduct between groups. In order to create new and efficiency methods, a meaning result help us deeply understand characteristic of reduct's definition in set-valued decision system and is a theory foundation to evaluate attribute reduction methods. Based on the results, further research is researching an alternative value of evaluating an decision reduct set's efficiency in set-valued decision system in order to complete problem.

References

[1] Guan Y. Y., Wang H. K., Set-valued information systems, *Information Sciences* 176, 2006, pp. 2507–2525.

[2] Kryszkiewicz M., Rough set approach to incomplete information systems, *Information Science*, Vol. 112, 1998, pp. 39-49.

[3] Lang G. M., Lia Q. G., Data compression of dynamic set-valued information systems, *CoRR abs/1209.6509*, 2012.

[4] Pawlak Z., Rough sets, *International Journal of Information and Computer Sciences*, 11(5), 1982, pp. 341-356.

[5] Pawlak Z., Rough sets: Theoretical Aspects of Reasoning About Data, Kluwer Aca-demic Publishers, 1991.

[6] Thi Thu Hien Phung, (2013). Generalized Discernibility Function based Attribute Reduction in Set-valued Decision Systems, *The Third World Congress on Information and Communication Technologies (WICT 2013), published by IEEE,* pp. 225-230.

[7] Qian Y. H., Dang C. Y., Liang J. Y., Tang D. W., Set-valued ordered information systems, *Information Sciences* 179, 2009, pp. 2809-2832.

[8] Sinh Hoa Nguyen, Thi Thu Hien Phung, "Efficient Algorithms for Attribute Reduction on Set-valued Decision Systems", *NXB Springer: Springer's Lecture Notes of Computer Science (LNCS) Series,* Volume 8170, 2013, pp 87-98.

[9] Skowron A., Rauszer C., The Discernibility Matrices and Functions in Information systems, Interlligent Decision Support, *Handbook of Applications and Advances of the Rough Sets Theory*, Kluwer, Dordrecht, 1992, pp. 331-362.

[10] Shifei D., Hao D., Research and Development of Attribute Reduction Algorithm Based on Rough Set,

IEEE, *CCDC2010*, 2010, pp. 648-653.
[11] Y. H. Qian Y. H. , Liang J. Y., On Dominance Relations in Disjunctive Set-Valued Ordered Information Systems, *International Journal of Information Technology & Decision Making,* Vol. 9, No. 1, 2010, pp. 9–33.
[12] Yuhua Qian, Chuangyin Dang, Jiye Liang, and Dawei Tang. 2009. Set-valued ordered information systems. Inf. Sci. 179, 16 (July 2009), 2809-2832.
[13] Wang C. Z., Chen D. G., Wuc C., Hu Q. H., Data compression with homomorphism in covering information systems, *International Journal of Approximate Reasoning* 52, 2011, pp. 519–525.
[14] Zhang J. B., Li T. R., Ruan D., Liu D., Rough sets based matrix approaches with dynamic attribute variation in set-valued information systems, *International Journal of Approximate Reasoning* 53, 2012, pp. 620–635.

Author Biographies

Thi Thu Hien Phung was born in Viet Nam on November 30, 1972. She is a lecturer in the Department of Information Technology at the University of Economic and Technical Industries, VietNam. Her area of research is Rough Set theory.

Thanh Dai Tran was born in Viet Nam on June 18, 1983. He is a lecturer in the Department of Information Technology at the University of Economic and Technical Industries, VietNam.