Soft Set Theoretic Approach for Dimensionality Reduction[1]

Tutut Herawan[1], Rozaida Ghazali[2], Mustafa Mat Deris[2]

[1]Department of Mathematics Education
Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[2]Faculty of Information Technology and Multimedia
Universiti Tun Hussein Onn Malaysia, Johor, Malaysia

tutut81@uad.ac.id, rozaida@uthm.edu.my, mmustafa@uthm.edu.my

Abstract

A reduct is a subset of attributes that are jointly sufficient and individually necessary for preserving a particular property of a given information system. The existing reduct approaches under soft set theory are still based on Boolean-valued information system. However, in the real applications, the data usually contain non-Boolean values. In this paper, an alternative approach for attribute reduction in multi-valued information system under soft set theory is presented. Based on the notion of multi-soft sets and AND operation, attribute reduction can be defined. It is shown that the reducts obtained are equivalent with Pawlak's rough reduction.

Keywords: *Information system; Reduct; Soft set theory.*

1. Introduction

In lots of data analysis applications, information and knowledge are stored and represented in an information table, where a set of objects is described by a set of attributes. To this, one practical problem is faced: for a particular property, whether all the attributes in the attribute set are always necessary to preserve this property [1]. Using the entire attribute set for describing the property is time-consuming, and the constructed rules may be difficult to understand, apply or verify. In order to deal with this problem, attribute reduction is required. The objective of reduction is to reduce the number of attributes, and at the same time, preserve the property of information. The theory of soft set [2] proposed, by Molodtsov 1999 is a new method for handling uncertain data. Soft sets are called (binary, basic, elementary) neighborhood systems [3]. The soft set is a mapping from parameter to the crisp subset of universe. From such case, we may see the structure of a soft set can classify the objects into two classes (yes/1 or no/0). This means that the "standard" soft set deals with a Boolean-valued information system. The theory of soft set has been applied to data analysis and decision support systems. A fundamental notion supporting such applications is the concept of reducts. The idea of dimensionality reductions under soft set theory have been proposed and compared, including the works of [4–8]. The restriction of those techniques is that they are applicable only for Boolean-valued information systems.

[1] An early version of this paper appeared in the Proceeding of International Conference, DTA 2009, held as Part of the Future Generation Information Technology Conference, FGIT 2009, Jeju Island, Korea, December 10-12, 2009, CCIS 64 Springer-Verlag, pp. 171–178, 2009.

However, in the theoretical and practical researches of soft sets, the situations are usually very complex and hence it may not suffice to represent data in the form of Boolean-valued information systems. In the real application, depending on the set of parameters, a given parameter may have different values (contain multiple grades). For example, the mathematics degree of student can be classified into three values; high, medium and low. In this situation, every parameter determines a partition of the universe which is contains more than two disjoint subsets. Unlike in Boolean-valued information systems, in multi-valued information systems, one cannot directly define the standard soft set. To this, we proposed the idea of *multi-soft sets* to deal multi-valued information systems. In this paper, we propose the idea of dimensionality reduction for multi-valued information systems under soft set theory. Three main contributions of this work are as follows: Firstly, we present the idea of multi-soft sets construction from a multi-valued information system, and AND and OR operations on multi-soft sets. Secondly, we present the applicability of the soft set theory for data reduction under multi-valued information system using multi-sets and AND operation. Lastly, we show that reducts obtained using soft set theory are equivalent to that rough set theory. Although some results are presented, a major part of this paper is devoted to revealing interconnection between reduction in multi-valued information systems under rough and soft set theories.

The rest of this paper is organized as follows. Section 2 describes related works of dimensionality reduction under soft set theory. Section 3 describes Information systems and set approximations. Section 4 describes fundamental soft set theory. Section 5 describes reduct in information systems using soft set theory. Finally, we conclude our works in section 6.

2. Related Works

The idea of reduct and decision making using soft set theory was firstly proposed by Maji *et al.* [4]. In [4], the application of soft set theory to a decision making problem with the help of Pawlak's rough mathematics was presented. The reduction approach presented is using Pawlak's rough reduction and a decision can be selected based on the maximal weighted value among objects related to the parameters. Chen *et al*. [5-6] presented the parameterization reduction of soft sets and its applications. They pointed out that the results of reduction proposed by Maji is incorrect and observed that the algorithms used to compute the soft set reduction and then to compute the choice value to select the optimal objects for the decision problem proposed by Maji are unreasonable. They also pointed out that the idea of reduct under rough set theory generally cannot be applied directly in reduct under soft set theory. The idea of Chen *et al.* for soft set reduction is only based on the optimal choice related to each object. However, the idea proposed by Chen is not error free, since the problems of the sub-optimal choice is not addressed. To this, Kong *et al.* [7] analyzed the problem of suboptimal choice and added parameter set of soft set. Then, they introduced the definition of normal parameter reduction in soft set theory to overcome the problems in Chen's model and described two new definitions, i.e. parameter important degree and soft decision partition and use them to analyze the algorithm of normal parameter reduction. With this approach, the optimal and sub-optimal choices are still preserved. Zou [8] proposed a new technique for decision making of soft set theory under incomplete information systems. The idea is based on the calculation of weighted-average of all possible choice values of object and the weight of each possible choice value is decided by the distribution of other objects. For fuzzy soft sets, incomplete data will be predicted based on the method of average probability. All those techniques are still based on Boolean information systems. As to this date, no researches have been done on dimensionality reduction in multi-valued information systems under soft set theory. Since every rough set [9] can be considered as soft set as presented in [10], thus, an alternative approach with potential for finding reduct in multi-valued information systems is using soft set theory. Still, it provides the same results for rough reduction [11−12].

3. Information Systems and Set Approximations

An information system is a 4-tuple (quadruple) $S=(U,A,V,f)$, where $U=\{u_1,u_2,u_3,\cdots,u_{|U|}\}$ is a non-empty finite set of objects, $A=\{a_1,a_2,a_3,\cdots,a_{|A|}\}$ is a non-empty finite set of attributes, $V=\bigcup_{a\in A}V_a$, V_a is the domain (value set) of attribute *a*, $f:U\times A\rightarrow V$ is an information function such that $f(u,a)\in V_a$, for every $(u,a)\in U\times A$, called information (knowledge) function.

An information system is also called a knowledge representation systems or an attribute-valued system. An information system can be intuitively expressed in terms of an information table (see Table 1).

Table 1. An information system

U	a_1	a_2	$\cdots$	a_k	$\cdots$	$a_{\|A\|}$
u_1	$f(u_1,a_1)$	$f(u_1,a_2)$	$\cdots$	$f(u_1,a_k)$	$\cdots$	$f(u_1,a_{\|A\|})$
u_2	$f(u_2,a_1)$	$f(u_2,a_2)$	$\cdots$	$f(u_2,a_k)$	$\cdots$	$f(u_2,a_{\|A\|})$
u_3	$f(u_3,a_1)$	$f(u_3,a_2)$	$\cdots$	$f(u_3,a_k)$	$\cdots$	$f(u_3,a_{\|A\|})$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$u_{\|U\|}$	$f(u_{\|U\|},a_1)$	$f(u_{\|U\|},a_2)$	$\cdots$	$f(u_{\|U\|},a_k)$	$\cdots$	$f(u_{\|U\|},a_{\|A\|})$

The complexity for computing an information system $S=(U,A,V,f)$ is $|U|\times|A|$ since there are $|U|\times|A|$ values of $f(u_i,a_j)$ to be computed, where $i=1,2,3,\cdots,|U|$ and $j=1,2,3,\cdots,|A|$. Note that t induces a set of maps $t=f(u,a):U\times A\rightarrow V$. Each map is a tuple $t_i=(f(u_i,a_1),f(u_i,a_2),f(u_i,a_3),\cdots,f(u_i,a_{|A|}))$, where where $i=1,2,3,\cdots,|U|$. Note that the tuple *t* is not necessarily associated with entity uniquely (see Example 6). In an information table, two distinct entities could have the same tuple representation (duplicated/redundant tuple), which is *not permissible* in relational databases. Thus, the concept of information systems is a generalization of the concept of relational databases. In many applications, there is an outcome of classification that is known. This a posteriori knowledge is expressed by one (or more) distinguished attribute called decision attribute; the process is known as supervised learning. An information system of this kind is called a decision system. A *decision system* [12] is an information system of the form $D=(U,A\cup\{d\},V,f)$, where $d\notin A$ is the decision attribute. The elements of *A* are called condition attributes.

The starting point of rough set approximations is the indiscernibility relation, which is generated by information about objects of interest. Two objects in an information system are called indiscernible (indistinguishable or similar) if they have the same feature.

Definition 1. (See [12].) *Let* $S=(U,A,V,f)$ *be an information system and let B be any subset of A. Two elements* $x,y\in U$ *are said to be B-indiscernible (indiscernible by the set of attribute* $B\subseteq A$ *in S) if and only if* $f(x,a)=f(y,a)$, *for every* $a\in B$.

Obviously, every subset of A induces unique indiscernibility relation. Notice that, an indiscernibility relation induced by the set of attribute B, denoted by $IND(B)$, is an equivalence relation. It is well known that, an equivalence relation induces unique partition. The partition of U induced by $IND(B)$ in $S=(U,A,V,f)$ denoted by U/B and the equivalence class in the partition U/B containing $x\in U$, denoted by $[x]_B$. The notions of lower and upper approximations of a set can be defined as follows.

Definition 2. (See [12].) *Let* $S=(U,A,V,f)$ *be an information system, let B be any subset of A and let X be any subset of U. The B-lower approximation of X, denoted by* $\underline{B}(X)$ *and B-upper approximations of X, denoted by* $\overline{B}(X)$, *are defined by*

$$\underline{B}(X)=\{x\in U \mid [x]_B\subseteq X\} \text{ and } \overline{B}(X)=\{x\in U \mid [x]_B\cap X\neq\phi\}, \text{ respectively.}$$

The notions of rough approximating of a set can be defined as follows:

Definition 3. *Let* $S=(U,A,V,f)$ *be an information system and let B be any subset of A. A rough approximation of a subset* $X\subseteq U$ *with respect to B is defined as a pair of lower and upper approximations of X, i.e.*

$$\left\langle \underline{B}(X),\overline{B}(X)\right\rangle.$$

Definition 4. *Let* $S=(U,A,V,f)$ *be an information system and let B be any subsets of A in information system S. Attribute* $b\in B$ *is called dispensable if*

$$U/(B-\{b\})=U/B.$$

Definition 5. *Let* $S=(U,A,V,f)$ *be an information system and let B be any subset of A in information system S. The subset* $B^*\subseteq B$ *is called reduct of B if* B^* *satisfies the following conditions:*

a. $U/B^*=U/B$
b. $U/(B^*-\{b\})\neq U/B, \quad \forall b\in B^*$

The core of B is defined as

$$\text{CORE}(B)=\bigcap \text{RED}(B),$$

where $\text{RED}(B)$ *is the set of all reducts of B.*

It is known that the problem of finding minimal reducts in information systems is NP-hard.

Example 6. For a simple example of rough reduction, we consider a small dataset derived from [13].

Table 1. An information system from [13]

U	a_1	a_2	a_3	a_4
1	low	bad	loss	small

2	low	good	loss	large
3	high	good	loss	medium
4	high	good	loss	medium
5	low	good	profit	large

Let $A = \{a_1, a_2, a_3, a_4\}$, then we have $U / A = \{\{1\}, \{2\}, \{3,4\}, \{5\}\}$. The reducts of A are $B = \{a_1, a_2, a_3\}$ and $C = \{a_3, a_4\}$, it causes

$$U / B = U / C = \{\{1\}, \{2\}, \{3,4\}, \{5\}\} = U / A .$$

The core is $\mathrm{CORE}(A) = B \cap C = \{a_3\}$.

The notion of soft sets and its fundamental operations are given in the following section. Much of the definition and examples are quoted directly from [2,14].

4. Soft Set Theory

Throughout this section *U* refers to an initial universe, *E* is a set of parameters, $P(U)$ is the power set of *U* and $A \subseteq E$.

Definition 7. (See [2].) *A pair* (F, A) *is called a soft set over U, where F is a mapping given by*

$$F : A \to P(U).$$

In other words, a soft set over *U* is a parameterized family of subsets of the universe *U*. For $\varepsilon \in A$, $F(\varepsilon)$ may be considered as the set of ε-elements of the soft set (F, A) or as the set of ε-approximate elements of the soft set. Clearly, a soft set is not a (crisp) set. To illustrate this idea, let we consider the following example.

Example 8. Let we consider a soft set (F, E) which describes the "attractiveness of houses" that Mr. X is considering to purchase.
Suppose that there are six houses in the universe *U* under consideration,

$$U = \{h_1, h_2, h_3, h_4, h_5, h_6\}$$

and

$$E = \{e_1, e_2, e_3, e_4, e_5\}$$

is a set of decision parameters, where e_1 stands for the parameters "expensive", e_2 stands for the parameters "beautiful", e_3 stands for the parameters "wooden", e_4 stands for the parameters "cheap", e_5 stands for the parameters "in the green surrounding".
Consider the mapping

$$F : E \to P(U),$$

given by "houses$(\cdot)$", where $(\cdot)$ is to be filled in by one of parameters $e \in E$.
Suppose that

$$F(e_1)=\{h_2,h_4\},\ F(e_2)=\{h_1,h_3\},\ F(e_3)=\{h_3,h_4,h_5\},\ F(e_4)=\{h_1,h_3,h_5\},\ F(e_5)=\{h_1\}.$$

Therefore, $F(e_1)$ means "houses (expensive)", whose functional value is the set $\{h_2,h_4\}$. Thus, we can view the soft set (F,E) as a collection of approximations as below

$$(F,E)=\left\{\begin{array}{c}\text{expensive houses}=\{h_2,h_4\},\\ \text{beautiful houses}=\{h_1,h_3\},\\ \text{wooden houses}=\{h_3,h_4,h_5\},\\ \text{cheap houses}=\{h_1,h_3,h_5\},\\ \text{in the green surrounding houses}=\{h_1\}\end{array}\right\}$$

Each approximation has two parts, a predicate *p* and an approximate value set *v*. For example, for the approximation "expensive houses$=\{h_2,h_4\}$", we have the predicate name of expensive houses and the approximate value set or value set if $\{h_2,h_4\}$.
Thus, a soft set (F,E) can be viewed as a collection of approximations below:

$$(F,E)=\{p_1=v_1,p_2=v_2,p_3=v_3,\cdots,p_n=v_n\}.$$

Table 4. Tabular representation of a soft set in the above example.

U	e_1	e_2	e_3	e_4	e_5
h_1	0	1	0	1	1
h_2	1	0	0	0	0
h_3	0	1	1	1	0
h_4	1	0	1	0	0
h_5	0	0	1	1	0
h_6	0	0	0	0	0

Definition 9. (See [14].) *The class of all value sets of a soft set* (F,E) *is called value-class of the soft set and is denoted by* $C_{(F,E)}$.

Clearly $C_{(F,E)} \subseteq P(U)$.

Proposition 10. *If* (F,E) *is a soft set over the universe U, then* (F,E) *is a Boolean-valued information system* $S=(U,A,V,f)$.

Proof. Let (F,E) be a soft set over the universe *U*, we define a mapping

$$F=\{f_1,f_2,\cdots,f_n\},$$

Where

$$f_1:U\to V_1 \text{ and } f_1(x)=\begin{cases}1, & x\in F(e_1)\\ 0, & x\notin F(e_1)\end{cases}$$

$$f_2 : U \to V_2 \text{ and } f_2(x) = \begin{cases} 1, & x \in F(e_2) \\ 0, & x \notin F(e_2) \end{cases}$$

$$\vdots$$

$$f_{|A|} : U \to V_{|A|} \text{ and } f_{|A|}(x) = \begin{cases} 1, & x \in F(e_{|A|}) \\ 0, & x \notin F(e_{|A|}) \end{cases}.$$

Thus, if $A = E$, $V = \bigcup_{e_i \in A} V_{e_i}$, where $V_{e_i} = \{0,1\}$, then a soft set (F, E) can be considered as a Boolean-valued information system $S = (U, A, V_{\{0,1\}}, f)$. □

From Proposition 10, it is easily to understand that a binary-valued information system can be represented as a soft set. Thus, we can make a one-to-one correspondence between (F, E) over U and $S = (U, A, V_{\{0,1\}}, f)$.

Proposition 11. *Every rough set can be considered as a soft set.*

Let $\langle \underline{B}(X), \overline{B}(X) \rangle$ be the rough approximating a subset $X \subseteq U$. We define a mapping $\underline{B}, \overline{B} : P(U) \to P(U)$ as in Definition 3. Thus every rough set $\langle \underline{B}(X), \overline{B}(X) \rangle$ can be considered a pair of two soft sets $(F, U) = \langle (\underline{B}, P(U)), (\overline{B}, P(U)) \rangle$. □

From the fact that every rough set can be considered as a soft set, in the following section we propose an alternative approach for dimensionality reduction in multi-valued information systems under soft set theory.

5. Reduct in Information Systems using Soft Set Theory

In this section we present the applicability of soft set theory for finding reducts. We show that the reducts obtained are equivalent to the rough reducts as in [11–12]. As for first step, we need a transformation from a multi-valued information system into multi-soft sets. In the multi-soft sets, we present the notion of AND and OR operations. For attribute reduction, we employ an AND operation and have managed to show that the reducts obtained are equivalent to rough reducts.

5.1 Multi-soft sets construction from multi-information systems

In this sub-section we discuss a decomposition of a multi-valued information system $S = (U, A, V, f)$ into $|A|$ number of binary-valued information systems. The decomposition of $S = (U, A, V, f)$ is based on decomposition of $A = \{a_1, a_2, \cdots, a_{|A|}\}$ into the disjoint-singleton attribute $\{a_1\}, \{a_2\}, \cdots, \{a_{|A|}\}$. At this stage, only complete information system is given the consideration. Let $S = (U, A, V, f)$ be an information system such that for every $a \in A$, $V_a = f(U, A)$ is a finite non-empty set and for every $u \in U$, $|f(u, a)| = 1$. For every a_i under i th-attribute consideration, $a_i \in A$ and $v \in V_a$, we define the map $a_v^i : U \to \{0,1\}$ such that $a_v^i(u) = 1$ if $f(u, a) = v$, otherwise $a_v^i(u) = 0$. The next result, we define a binary-valued information system as a quadruple $S^i = (U, a_i, V_{\{0,1\}}, f)$. The information systems

$S^i = \left(U, a_i, V_{\{0,1\}}, f\right)$, $i = 1,2,\cdots,|A|$ is referred to as a decomposition of a multi-valued information system $S = (U, A, V, f)$ into $|A|$ binary-valued information systems, as depicted in Figure 1. Every information system $S^i = \left(U, a_i, V_{ai}, f\right)$, $i = 1,2,\cdots,|A|$ is a deterministic information system since for every $a \in A$ and for every $u \in U$, $|f(u,a)| = 1$ such that the structure of a multi-valued information system and $|A|$ number of binary-valued information systems give the same value of attribute related to objects.

U	a_1	a_2	$\cdots$	a_k	$\cdots$	$a_{\|A\|}$
u_1	$f(u_1,a_1)$	$f(u_1,a_2)$	$\cdots$	$f(u_1,a_k)$	$\cdots$	$f(u_1,a_{\|A\|})$
u_2	$f(u_2,a_1)$	$f(u_2,a_2)$	$\cdots$	$f(u_2,a_k)$	$\cdots$	$f(u_2,a_{\|A\|})$
u_3	$f(u_3,a_1)$	$f(u_3,a_2)$	$\cdots$	$f(u_3,a_k)$	$\cdots$	$f(u_3,a_{\|A\|})$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$u_{\|U\|}$	$f(u_{\|U\|},a_1)$	$f(u_{\|U\|},a_2)$	$\cdots$	$f(u_{\|U\|},a_k)$	$\cdots$	$f(u_{\|U\|},a_{\|A\|})$

Binary-valued information system-1

U	a_1					
	Va_{11}	Va_{12}	$\cdots$	Va_{1k}	$\cdots$	Va_{1n}
u_1	0	1	$\cdots$	0	$\cdots$	0
u_2	0	0	$\cdots$	1	$\cdots$	0
u_3	0	1	$\cdots$	0	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$u_{\|U\|}$	1	0	$\cdots$	0	$\cdots$	0

$\vdots$

Binary-valued information system-$|A|$

U	$a_{\|A\|}$					
	$Va_{\|A\|1}$	$Va_{\|A\|2}$	$\cdots$	$Va_{\|A\|k}$	$\cdots$	$Va_{\|A\|n}$
u_1	0	0	$\cdots$	1	$\cdots$	0
u_2	0	0	$\cdots$	0	$\cdots$	1
u_3	0	1	$\cdots$	0	$\cdots$	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$u_{\|U\|}$	1	0	$\cdots$	0	$\cdots$	0

Figure 1. A decomposition of a multi-valued information system

Based on the notion of a decomposition of a multi-valued information system in the previous sub-section, in this sub-section we present the notion of multi-soft set representing multi-valued information systems. Let $S = (U, A, V, f)$ be a multi-valued information system and $S^i = \left(U, a_i, V_{ai}, f\right)$, $i = 1,2,\cdots,|A|$ be the $|A|$ binary-valued information systems. From Proposition 10, we have

$$S=(U,A,V,f)=\begin{cases} S^1=(U,a_1,V_{\{0,1\}},f) & \Leftrightarrow (F,a_1) \\ S^2=(U,a_2,V_{\{0,1\}},f) & \Leftrightarrow (F,a_2) \\ \quad\vdots & \quad\vdots \quad \vdots \\ S^{|A|}=(U,a_{|A|},V_{\{0,1\}},f) & \Leftrightarrow (F,a_{|A|}) \end{cases}$$
$$=((F,a_1),(F,a_2),\cdots,(F,a_{|A|}))$$

We define $(F,E)=((F,a_1),(F,a_2),\cdots,(F,a_{|A|}))$ as a multi-soft set over universe U representing a multi-valued information system $S=(U,A,V,f)$.

Example 11. The multi Boolean information systems representing Table 4 is given below

U	a_1	
	low	**high**
1	1	0
2	1	0
3	0	1
4	0	1
5	1	0

U	a_2	
	bad	**good**
1	1	0
2	0	1
3	0	1
4	0	1
5	0	1

U	a_3	
	loss	**profit**
1	1	0
2	1	0
3	1	0
4	1	0
5	0	1

U	a_4		
	small	**large**	**medium**
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	1
5	0	1	0

Figure 2. A decomposition of a multi-valued information system

From Figure 2, we have the following corresponding soft sets

$$(F, a_1) = \{\{\text{low} = 1,2,5\}, \{\text{high} = 3,4\}\},$$
$$(F, a_2) = \{\{\text{bad} = 1\}, \{\text{good} = 2,3,4,5\}\},$$
$$(F, a_3) = \{\{\text{loss} = 1,2,3,4\}, \{\text{profit} = 5\}\},$$
$$(F, a_4) = \{\{\text{small} = 1\}, \{\text{large} = 2,5\}, \{\text{medium} = 3,4\}\}.$$

Thus, the multi-soft set representing Table 2 is

$$(F, A) = ((F, a_1), (F, a_2), (F, a_3), (F, a_4)) = \begin{pmatrix} \{\{\text{low} = 1,2,5\}, \{\text{high} = 3,4\}\}, \{\{\text{bad} = 1\}, \{\text{good} = 2,3,4,5\}\}, \\ \{\{\text{loss} = 1,2,3,4\}, \{\text{profit} = 5\}\}, \{\{\text{small} = 1\}, \{\text{large} = 2,5\}, \{\text{medium} = 3,4\}\} \end{pmatrix}$$

5.2 AND and OR operations in multi-soft sets

The notions of AND and OR operations in multi-soft sets are given below.

Definition 12. *Let* $(F, E) = ((F, a_i) : i = 1,2, \cdots, |A|)$ *be a multi-soft set over U representing a multi-valued information system* $S = (U, A, V, f)$. *The* AND *operation between* (F, a_i) *and* (F, a_j) *is defined as*

$$(F, a_i)\text{AND}(F, a_j) = (F, a_i \times a_j),$$

where

$$G(Va_i, Va_j) = F(Va_i) \cap F(Va_j),\ \forall (Va_i, Va_j) \in a_i \times a_j,\ for\ 1 \le i, j \le |A|.$$

Example 13. From Example11, let two soft-sets

$$(F, a_1) = \{\{\text{low} = 1,2,5\}, \{\text{high} = 3,4\}\}$$

and

$$(F, a_2) = \{\{\text{bad} = 1\}, \{\text{good} = 2,3,4,5\}\}.$$

Then, we have

$$(F, a_1)\text{AND}(F, a_2) = (F, a_1 \times a_3)$$
$$= ((\text{low, bad}) = \{1\}, (\text{low, good}) = \{2,5\}, (\text{high, bad}) = \phi, (\text{high, good}) = \{3,4\}).$$

Definition 14. *Let* $(F, E) = ((F, a_i) : i = 1,2, \cdots, |A|)$ *be a multi-soft set over U representing a multi-valued information system* $S = (U, A, V, f)$. *The* OR *operation between* (F, a_i) *and* (F, a_j) *is defined as*

$$(F, a_i)\text{OR}(F, a_j) = (F, a_i \times a_j),$$

where

$$G(Va_i, Va_j) = F(Va_i) \cup F(Va_j),\ \forall (Va_i, Va_j) \in a_i \times a_j,\ for\ 1 \le i, j \le |A|.$$

Example 15. From Example 5, let two soft-sets

$$(F, a_2) = \{\{\text{bad} = 1\}, \{\text{good} = 2,3,4,5\}\},$$

And

$$(F,a_3)=\{\{\text{loss}=1,2,3,4\},\{\text{profit}=5\}\},$$

Then we have

$$(F,a_2)\text{OR}(F,a_3)=(F,a_2\times a_3)= \\ =\begin{pmatrix}(\text{bad},\text{loss})=\{1,2,3,4\},(\text{bad},\text{profit})=\{1,5\}, \\ (\text{good},\text{loss})=\{1,2,3,4,5\},(\text{good},\text{profit})=\{2,3,4,5\}\end{pmatrix}$$

5.3 Attribute reduction

In this section we propose the idea of attributes reduction under soft set theory. The proposed approach is based on AND operation in multi-soft sets as described in the previous section.

Definition 16. *Let* $(F,A)=((F,a_i):i=1,2,\cdots,|A|)$ *be a multi-soft set over U representing a multi-valued information system* $S=(U,A,V,f)$. *A set of attributes* $B\subseteq A$ *is called a reduct for A if* $C_{F(b_1\times\cdots\times b_{|B|})}=C_{F(a_1\times\cdots\times a_{|A|})}$.

Example 17. From Example 5, let two multi soft-sets

$$(F,\{a_1,a_2,a_3\})\text{ and }(F,\{a_3,a_4\})$$

a. For $(F,\{a_1,a_2,a_3\})$, where $(F,a_1)=\{\{\text{low}=1,2,5\},\{\text{high}=3,4\}\}$, $(F,a_2)=\{\{\text{bad}=1\},\{\text{good}=2,3,4,5\}\}$ and $(F,a_3)=\{\{\text{loss}=1,2,3,4\},\{\text{profit}=5\}\}$.

Then, we have

$$(F,a_1)\text{AND}(F,a_2)\text{AND}(F,a_3)=(F,a_1\times a_2\times a_3) \\ =\begin{pmatrix}(\text{low},\text{bad},\text{loss})=\{1\},(\text{low},\text{bad},\text{profit})=\{5\}, \\ (\text{low},\text{good},\text{loss})=\{2\},(\text{low},\text{good},\text{profit})=\{5\}, \\ (\text{high},\text{bad},\text{loss})=\phi,(\text{high},\text{bad},\text{profit})=\phi, \\ (\text{high},\text{good},\text{loss})=\{3,4\},(\text{high},\text{good},\text{profit})=\{5\}\end{pmatrix}$$

Notice that,

$$C_{F(a_1\times a_2\times a_3)}=\{\{1\},\{2\},\{3,4\},\{5\}\}. \quad (1)$$

b. For $(F,\{a_3,a_4\})$, where $(F,a_3)=\{\{\text{loss}=1,2,3,4\},\{\text{profit}=5\}\}$ and $(F,a_4)=\{\{\text{small}=1\},\{\text{large}=2,5\},\{\text{medium}=3,4\}\}$,

Then we have

$$(F,a_3)\text{AND}(F,a_4)=(F,a_3\times a_4) \\ =\begin{pmatrix}(\text{loss},\text{small})=\{1\},(\text{loss},\text{large})=\{2\},(\text{loss},\text{medium})=\{3,4\}, \\ (\text{profit},\text{small})=\phi,(\text{profit},\text{large})=\{5\},(\text{profit},\text{medium})=\phi\end{pmatrix}$$

Notice that,

$$C_{F(a_3 \times a_4)} = \{\{1\}, \{2\}, \{3,4\}, \{5\}\}. \qquad (2)$$

From (1) and (2), we have $\{a_1, a_2, a_3\}$ and $\{a_3, a_4\}$ are reducts of A.

6. Conclusion

The existing reduct approaches under soft set theory are still based on Boolean-valued information system. For the real applications, the data usually contain non-Boolean valued. In this paper, an alternative approach for attribute reductions in multi-valued information systems under soft set theory has been presented. In the proposed approach, the notion of multi-soft set is used to represent multi-valued information systems. The AND operation is used in multi-soft sets to present the notion of attribute reduction. It is founded that the obtained reducts are equivalent to the rough reducts.

Acknowledgement

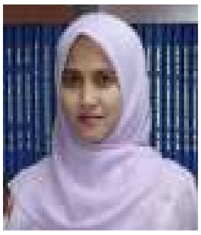
This work was supported by the grant of Universiti Tun Hussein Onn Malaysia.

References

[1] Zhao, Y., Luo, F., Wong, S.K.M. and Yao, Y.Y. “A general definition of an attribute reduct”. Proceeding of Second International Conference on Rough Sets and Knowledge Technology, RSKT 2007, LNAI 4481, 101–108.
[2] Molodtsov, D. “Soft set theory-first results”. Computers and Mathematics with Applications. 37, 1999, 19–31.
[3] Yao, Y.Y. “Relational interpretations of neighbourhood operators and rough set approximation operators, Information Sciences, 111, 1998, 239–259.
[4] Maji, P.K., Roy, A.R., and Biswas, R. “An application of soft sets in a decision making problem”. Compututer and Mathematics with Application, 44, 2002.
[5] Chen, D., Tsang, E.C.C., Yeung, D.S., and Wang, X. “Some notes on the parameterization reduction of soft sets”. Proceeding of International Conference on Machine Learning and Cybernetics, 3, 2003, IEEE Press, 1442–1445.
[6] Chen, D., Tsang, E.C.C., Yeung, D.S., and Wang, X. “The Parameterization Reduction of Soft Sets and its Applications”. Computers and Mathematics with Applications, 49, 2005, 757–763.
[7] Kong, Z., Gao, L., Wang, L., and Li, S. “The normal parameter reduction of soft sets and its algorithm”. Computers and Mathematics with Applications 56, 2008, 3029–3037.
[8] Zou, Y. and Xiao, Z. “Data analysis approaches of soft sets under incomplete information”. Knowledge Based Systems, 21, 2008, 941–945.
[9] Pawlak, Z. “Rough sets”. International Journal of Computer and Information Science. 11, 1982, 341–356.

[10] Herawan, T. and Mustafa M.D. "A direct proof of every rough set is a soft set". Proceeding of International Conference AMS 2009, IEEE Press, 119-124.
[11] Pawlak, Z. "Rough sets: A theoretical aspect of reasoning about data". Kluwer Academic Publisher, 1991.
[12] Pawlak, Z. and Skowron, A. "Rudiments of rough sets". Information Sciences, 177, 1, 2007, 3–27.
[13] Pawlak, Z. "Rough classification". International journal of human computer studies, 51, 1999, 369-383.
[14] Maji, P.K., Biswas, R., and Roy, A.R. "Soft set theory". Computers and Mathematics with Applications. 45, 2003, 555–562.

Tutut Herawan
He is a Ph.D. candidate in Data Mining at Universiti Tun Hussein Onn Malaysia (UTHM). His research area includes Data Mining, KDD and Real Analysis.



Rozaida Ghazali
She received her B.Sc. (Hons) degree in Computer Science from Universiti Sains Malaysia, and M.Sc. degree in Computer Science from Universiti Teknologi Malaysia. She obtained her Ph.D. degree in Higher Order Neural Networks at Liverpool John Moores University, UK. She is currently a teaching staff at Faculty of Information technology and Multimedia, Universiti Tun Hussein Onn Malaysia (UTHM). Her research area includes neural networks, fuzzy logic, financial time series prediction and physical time series forecasting.

Mustafa Mat Deris
He received the B.Sc. from University Putra Malaysia, M.Sc. from University of Bradford, England and Ph.D. from University Putra Malaysia. He is a professor of computer science in the Faculty of Information Technology and Multimedia, UTHM, Malaysia. His research interests include distributed databases, data grid, database performance issues and data mining. He has published more than 80 papers in journals and conference proceedings. He was appointed as one of editorial board members for International Journal of Information Technology, World Enformatika Society, a reviewer of a special issue on International Journal of Parallel and Distributed Databases, Elsevier, 2004, a special issue on International Journal of Cluster Computing, Kluwer, 2004, IEEE conference on Cluster and Grid Computing, held in Chicago, April, 2004, and Malaysian Journal of Computer Science. He has served as a program committee member for numerous international conferences/workshops including Grid and Peer-to-Peer Computing, (GP2P 2005, 2006), Autonomic Distributed Data and Storage Systems Management (ADSM 2005, 2006), WSEAS, International Association of Science and Technology, IASTED on Database, etc.