International Journal of Pure and Applied Mathematics
Volume 27 No. 1 2006, 39-53

NEW APPROACHES FOR DATA REDUCTION

A.S. Salama[1 §], H.M. Abu-Donia[2]

[1]Department of Mathematics
Faculty of Science
Tanta University
Tanta, EGYPT
e-mail: amgadsalama2003@yahoo.com

[2]Department of Mathematics
Faculty of Science
Zagazig University
Zagazig, EGYPT
e-mail: donia_1000@yahoo.com

Abstract: In this paper, we studied some topological properties of information systems and we introduced three new approaches for data reduction. Topological approach for data reduction is a new method to deal with general types of relations. The reducts of an information systems has here some orders (first order, second order, and so on) and also the core. The second approach depends on the comparing the values of each subset of the set of condition attributes with the decision attribute. The evaluation of reduct and the core by the second approach is a quick and efficiently method for data reduction than the classical methods. The last approach depends on the notion of topological covering.

AMS Subject Classification: 93C57
Key Words: information systems, topological spaces, and data reduction

1. Introduction

Information systems, introduced by Z. Pawlak in (1982) [12, 14, 15, 16] are excellent tools to handle a granularity of data. It may be used to describe

Received: December 24, 2005

§Correspondence author

dependencies between attributes, to evaluate significance of attributes, and to deal with inconsistent data, to name just a few possible uses out of many ways of analysis information systems to real world problems. Most important is an approach to handling imperfect data.

The calculus of these systems is based on an objective viewpoint of the world; i.e., all computations of the calculus are based on existing data characteristics. Many other ways of handling imperfect data are subjective, world situations by experts.

The notion of indiscernibility, the main idea of rough set theory, is closely related to data granularity. The notion of indiscernibility may be introduced in two different ways. First, it may discuss in the more general but less intuitive form as the fair of the universe. Second, it may also be presented in the form of a table, called an information table or an information system. We will use the latter approach in this paper because it is more application oriented.

The rough set theory is based also on complete information systems [10, 11, 13]. It classifies objects using upper – approximation and lower – approximation defined on the indiscernibility relation. In order to process incomplete information systems, the rough set theory needs to be extended, especially, the indiscernibility relation needs to be extended to some in-equivalent relations there are several extensions for the indiscernibility relation at present [7, 9], such as tolerance relations non-symmetric similarity relations and complementarily relations.

2. Topological Properties of Information Systems

Studying the topological applications of information systems appeared in [5, 7, 9]. An information system can be defined by a quadruple, $S = (Ob, At, \{V_a : a \in At\}, f_a)$ where:

— Ob is a finite non-empty set of objects,

— At is a finite non-empty set of attributes,

— V_a is a finite non-empty set of values of $a \in At$,

— $f_a : Ob \longrightarrow P(V_a)$ is an information function.

For any object $x \in Ob$ when $f_a(x) \in V_a$ for all $a \in At$, the information system S in this case called single value information system (Pawlak systems). On the other hand, when $f_a(x) \in P(V_a)$ it is called set value information system. An information system S is complete if for all $a \in At$ and for all $x \in Ob$, $f_a(x) \neq \phi$.

Special type of information systems, which is called a nominal information

Ob	Inf_1	Inf_2	Inf_3
x_1	a	d	i
x_2	a	e	h
x_3	b	e	g
x_4	c	f	g
x_5	b	d	i

Table 2.1:

system and it is defined as: $S = (Ob, \sum, \{\mathrm{Inf}_n\}_{n\in M})$ where:

— M is a finite set of positive integers,

— Ob is a finite non-empty set of objects,

— $\sum$ is a finite non - empty set of alphabet of information symbols,

— $\mathrm{Inf}_n : Ob \longrightarrow \sum$ for $n \in M$ is information function satisfying the condition:

For every $n, m \in M$ and $w \in \sum$, if $n > m$, then $\mathrm{Inf}_n^{-1}(\mathrm{Inf}_n(X_w^m)) = X_w^m$, where $X_w^m = \mathrm{Inf}_n^{-1}(w) = \{x \in Ob : \mathrm{Inf}_m^{-1}(x) = w\}$. The set X_w^m is called information set with information w on the level m. For any subsets $A \subseteq M$ and $B \subseteq \sum$, the set X_B^A, where $X_B^A = \{x \in Ob : \mathrm{Inf}_n^{-1}(x) \in B, n \in A\}$ is called an information set with information set B on the level $n \in A$.

Example 2.1. Consider $Ob = \{x_1, x_2, x_3, x_4, x_5\}$ be a set of objects and let $M = \{1, 2, 3\}$ and $\sum = \{a, b, c, d, e, f, g, h, i\}$ is the set of alphabet of information symbols. Then Table 2.1 represents a nominal information system.

According to Table 2.1 we have:

$$X_a^1 = \{x_1, x_2\},\ X_b^1 = \{x_3, x_5\},\ X_c^1 = \{x_4\},$$

$$X_d^2 = \{x_1, x_5\},\ X_e^2 = \{x_2, x_3\},\ X_f^2 = \{x_4\},$$

$$X_i^3 = \{x_1, x_5\},\ X_h^3 = \{x_2\},\ X_g^3 = \{x_3, x_4\},$$

$$X_{\{d,i\}}^{\{2,3\}} = \{x_1, x_5\},\ X_{\{e,h\}}^{\{2,3\}} = \{x_2\},\ X_{\{e,g\}}^{\{2,3\}} = \{x_3\},\ X_{\{f,g\}}^{\{2,3\}} = \{x_4\}.$$

Proposition 2.1. *The families $\{X_w^n\}$, $n \in M$, $w \in \sum$ are partitions of objects in any nominal information system $S = (Ob, \sum, \{\mathrm{Inf}_n\}_{n\in M})$.*

Proof. The families $\{X_w^n\}$, $n \in M$, $w \in \sum$ are the families of equivalence classes of the equivalence relation E_n defined on Ob by: $(x, y) \in E_n, n \in M$ iff $\mathrm{Inf}_n(x) = \mathrm{Inf}_n(y)$. □

Proposition 2.2. *The families $\{X_w^n\}$, $n \in M$, $w \in \sum$ are topological bases.*

Proof. From Proposition 2.1. □

Lemma 2.1. *If β is a base for a topological space (U, τ), where β is a partition of U, then for every subset $X \subseteq U$:*

(i) $\mathrm{Int}_\tau(X) = \bigcup\{B \in \beta : B \subseteq X\}$.

(ii) $\mathrm{Cl}_\tau(X) = \bigcup\{B \in \beta : B \cap X \neq \phi\}$.

Proof. Only we prove (ii) because (i) is trivial. Let $x \in \mathrm{Cl}_\tau(X)$, then for every open set G containing x, $X \cap G \neq \phi$. But $G = \bigcup_{B \in \beta} B$, then there exists $B_0 \in \beta$ such that $x \in B_0 \subseteq G$.

But B_0 is an open set containing x, hence $B_0 \cap X \neq \phi$ and $x \in \bigcup\{B \in \beta : B \cap X \neq \phi\}$.

Conversely, if $x \in \bigcup\{B \in \beta : B \cap X \neq \phi\}$ and G is an open set containing x and β is a partition of U, $x \in U$, then x belongs to only one element of β say $x \in B_0$. Then must $B_0 \subseteq G$, i.e., $x \in B_0 \subseteq G$ but $B_0 \cap X \neq \phi$, hence $G \cap X \neq \phi$. Then $x \in \mathrm{Cl}_\tau(X)$.

Let $S = (Ob, \sum, \{\mathrm{Inf}_n\}_{n \in M})$ be a nominal information system. For any subset Y of Ob and $A \subseteq M$ we define: $L(X) = \bigcup\{X_B^A : X_B^A \subseteq Y\}$ and $U(X) = \bigcup\{X_B^A : X_B^A \cap Y \neq \phi\}$.

For a given subset B of $\sum$, $L(X)$ and $U(X)$ are called the lower and the upper approximations of Y in S respectively. According to Example 2.1, if $A = \{1\}$ and $B = \{a, b, c\}$ then $L(Y) = \{x_4\}$ and $U(Y) = \{x_1, x_2, x_4\}$ for the subset $Y = \{x_1, x_4\}$. □

Theorem 2.1. *Let $S = (Ob, \sum, \{\mathrm{Inf}_n\}_{n \in M})$ be a nominal information system. For any $Y \subseteq Ob, A \subseteq M$ and $B \subseteq \sum$ we have $\mathrm{Int}_\tau(Y) = L(Y)$ and $\mathrm{Cl}_\tau(Y) = U(Y)$, where τ is the topology generated by the base $\{X_B^A\}$.*

Proof. From Lemma 2.1. □

Theorem 2.2. *Topologies generated by the bases $\{X_B^A\}$ are all quasi-discrete topological spaces.*

Proof. A set X in a quasi-discrete topological space is open iff it is closed. If X is open then $X = \bigcup\{X_B^A : A \subset M, B \subseteq \sum\}$. Hence $X = U - \bigcup\{X_B^A : B' \subseteq \sum - B\}$ so X is closed. Also if X is closed then $X = U - \bigcup\{X_B^A : B \subseteq \sum\}$ for some $B \subseteq \sum$. Hence $X = \bigcup\{X_B^A : B' \subseteq \sum - B\}$ and X is open. □

According to Example 2.1, the following are bases for topologies on Ob:

$\beta_1 = \{X_a^1, X_b^1, X_c^1\} = \{\{x_1, x_2\}, \{x_3, x_5\}, \{x_4\}\}$,

$\beta_2 = \{X_d^2, X_e^2, X_f^2\} = \{\{x_1, x_5\}, \{x_2, x_3\}, \{x_4\}\}$,

$\beta_3 = \{X_i^3, X_h^3, X_g^3\} = \{\{x_1, x_5\}, \{x_2\}, \{x_3, x_4\}\}$,

$\beta_4 = \{X_{\{d,i\}}^{\{2,3\}}, X_{\{e,h\}}^{\{2,3\}}, X_{\{e,g\}}^{\{2,3\}}, X_{\{f,g\}}^{\{2,3\}}\} = \{\{x_1, x_5\}, \{x_2\}, \{x_3\}, \{x_3\}\}$.

If $Y = \{x_2, x_3, x_5\}$ be a subset of Ob, then with respect to the base β_3 we have $L(Y) = \{x_3\}$, $U(Y) = \{x_1, x_2, x_3, x_4, x_5\}$, $\text{Int}_{\tau_{\beta_3}}(Y) = \{x_3\}$ and $\text{Cl}_{\tau_{\beta_3}}(Y) = \{x_1, x_2, x_3, x_4, x_5\}$.

Let $S = (Ob, \sum, \{\text{Inf}_n\}_{n\in M})$ be a nominal information system. For any two levels $n, m \in M$ and any values $w, w^{'} \in \sum$ let $\{X^n_w\}$ and $\{X^m_{w^{'}}\}$ be two partitions of the set of objects Ob defined by the equivalence relations Inf_n and Inf_m respectively. Then we say that the partition $\{X^m_{w^{'}}\}$ depends on the partition $\{X^n_w\}$ denoted $\{X^n_w\} \leq \{X^m_{w^{'}}\}$ if and only if: $X^m_{w^{'}} = \bigcup_n X^n_w$ For all $X^m_{w^{'}} \in \{X^m_{w^{'}}\}$.

Theorem 2.3. *Let τ_n and τ_m be the topologies induced by the partitions $\{X^n_w\}$ and $\{X^m_{w^{'}}\}$ respectively. Then $\{X^n_w\} \leq \{X^m_{w^{'}}\}$ iff $\tau_m \subseteq \tau_n$.*

Proof. Let $G \in \tau_m$ be an open set, then $G = \bigcup_m X^m_{w^{'}}$, $X^m_{w^{'}} \subseteq G$ for some $m \in M$, where $\{X^m_{w^{'}}\}$ is a base of τ_m. But $X^m_{w^{'}} = \bigcup_n X^n_w$, hence $G = \bigcup_m \bigcup_n X^n_w$, i.e., $G = \bigcup_{max\{n,m\}} X^n_w$ which implies that $G \in \tau_n$, hence $\tau_m \subseteq \tau_n$. Conversely, if $\tau_m \subseteq \tau_n$, then for every $G \in \tau_m$ also implies $G \in \tau_n$, hence $G = \bigcup_m X^m_{w^{'}} = \bigcup_n X^n_w$ then there exist n_0 such that $X^m_{w^{'}} = \bigcup_{n_0} X^{n_0}_w$. Hence $\{X^n_w\} \leq \{X^m_{w^{'}}\}$. □

Example 2.2. Consider the partitions $\beta_1 = \{\{x_1, x_2\}, \{x_3\}, \{x_4\}\}$ and $\beta_2 = \{\{x_1, x_2\}, \{x_3, x_4\}\}$ of the set $U = \{x_1, x_2, x_3, x_4\}$. Then $\beta_1 \leq \beta_2$ and $\tau_2 \subseteq \tau_1$, where:

$\tau_1 = \{U, \phi, \{x_3\}, \{x_4\}, \{x_1, x_2\}, \{x_3, x_4\}, \{x_1, x_2, x_3\}, \{x_1, x_2, x_4\}\}$, and

$\tau_2 = \{U, \phi, \{x_1, x_2\}, \{x_3, x_4\}\}$ are the topologies generated by β_1 and β_1 respectively.

For any topological space (U, τ) , we define the equivalence relation $E(\tau)$ on the set U by: $(x, y) \in E(\tau)$ iff $\text{Cl}_\tau(\{x\}) = \text{Cl}_\tau(\{y\})$ $x, y \in U$. The set of all equivalence classes of $E(\tau)$ is denoted by $U/E(\tau)$.

Theorem 2.4. *Let $S = (U, \sum, \{\text{Inf}_n\}_{n\in M})$ be a nominal information system and let τ_n be the topology generated by the base $\beta_n = \{X^n_w : n \in M, w \in \sum\}$. If (U, τ) be any quasi-discrete topological space has $U/E(\tau)$ as a base. Then $\tau_n = \tau$ iff for all $x \in X^n_w$ there exists $B \in U/E(\tau)$ such that $x \in B$.*

Proof. If for all $x \in X^n_w$ there exists $B \in U/E(\tau)$ and $x \in B$, then $X^n_w = B$ hence $B \in U/E(\tau)$ and $\tau_n = \tau$.

Conversely, if $\tau_n = \tau$, and τ_n is a quasi-discrete topological space generated by β_n, then τ_n and τ have the same base i.e., for all $X^n_w \in \beta_n$ there exist $B \in U/E(\tau)$ such that $X^n_w = B$, hence for all $x \in X^n_w$, there exist $B \in U/E(\tau)$ such that $x \in B$. □

Lemma 2.2. (see [9]) *For any topology τ on a set U, and for all $x, y \in U$,* *if* $x \in \mathrm{Cl}_\tau(\{y\})$ *and* $y \in \mathrm{Cl}_\tau(\{x\})$ *then* $\mathrm{Cl}_\tau(\{x\}) = \mathrm{Cl}_\tau(\{y\})$.

Lemma 2.3. (see [9]) *If* τ *is a quasi-discrete topology on a set* U, *then* $y \in \mathrm{Cl}_\tau(\{x\})$ *implies* $x \in \mathrm{Cl}_\tau(\{y\})$ *for all* $x, y \in U$.

Lemma 2.4. (see [9]) *If* τ *is a quasi-discrete topology on a set* U, *then the family* $\{\mathrm{Cl}_\tau(\{x\}) : x \in U\}$ *is a partition of* U.

Proposition 2.3. *Let* τ *be the topology induced by the partition* $\beta_n = \{X_w^n : n \in M, w \in \sum\}$ *on the set* Ob, *where* $S = (Ob, \sum, \{\mathrm{Inf}_n\}_{n\in M})$ *be a nominal information system. Then* $\beta_n = Ob/E(\tau)$.

Proof.

$$\begin{aligned} x \in B, B \in \beta_n \quad &\Leftrightarrow \quad x \in \mathrm{Cl}_\tau(B) = \bigcup_{y\in B} \mathrm{Cl}_\tau(\{y\}) \\ &\Leftrightarrow \quad y_0 \in B \text{ and } x \in \mathrm{Cl}_\tau(\{y_0\}) \\ &\Leftrightarrow \quad \mathrm{Cl}_\tau(\{x\}) = \mathrm{Cl}_\tau(\{y_0\}) \text{ (Lemma 2.2)} \\ &\Leftrightarrow \quad (x, y_0) \in E(\tau) \\ &\Leftrightarrow \quad \exists A \in Ob/E(\tau) \text{ suchthat } x \in A \\ &\Leftrightarrow \quad \beta_n = Ob/E(\tau)\,. \end{aligned}$$

For any nominal information system $S = (Ob, \sum, \{\mathrm{Inf}_n\}_{n\in M})$ and for all $n \in M$ we define the partition $Ob/E(\tau_{ind}) = \bigcap_{n\in M} \{Ob/E(\tau_n)\}$. □

Theorem 2.5. *For any nominal information system* $S = (Ob, \sum, \{\mathrm{Inf}_n\}_{n\in M})$, *then* $\tau_n \subseteq \tau_{ind}$ *where* τ_n *and* τ_{ind} *are the topologies generated by the partitions* $Ob/E(\tau_n)$ *and* $Ob/E(\tau_{ind})$ *respectively.*

Proof. Since $Ob/E(\tau_{ind}) \leq Ob/E(\tau_n)$ for all $n \in M$ then $\tau_n \subseteq \tau_{ind}$ (see Theorem 2.3). □

Example 2.3. Consider the topological space (U, τ) where $U = \{x_1, x_2, x_3, x_4\}$ and $\beta = \{\{x_1\}, \{x_2, x_3\}, \{x_4\}\}$ is the base of τ, then τ is a quasi-discrete topology and $\mathrm{Cl}_\tau(\{x_1\}) = \{x_1\}, \mathrm{Cl}_\tau(\{x_2\}) = \{x_2, x_3\}, \mathrm{Cl}_\tau(\{x_3\}) = \{x_2, x_3\}, \mathrm{Cl}_\tau(\{x_4\}) = \{x_4\}$.

Then $U/E(\tau) = \{\{x_1\}, \{x_2, x_3\}, \{x_4\}\} = \beta$.

3. Topological Reduction of Data

By reduction we mean if we can remove some data from the data table given in our information system preserving its basic properties. To express this idea more precisely, let $S = (Ob, At, \{V_a : a \in At\}, f_a)$ be an information system (numerical system). Let r be a positive real, for each object $x \in Ob$ and for $a \in At$, $N_a(x, r)$ is the $a-$neighborhood of x and defined by:

$$N_a(x, r) = \{y \in Ob :\mid f_a(x) - f_a(x) \mid\leq r\}.$$

For any subset B of At , the $B-$neighborhood of x is defined by:

$$N_B(x,r) = \{y \in Ob :\mid f_a(x) - f_a(x) \mid \leq r \; \forall a \in B\}.$$

For any subset X of Ob, we define two mappings: $\mathrm{Int}, \mathrm{Cl} : P(Ob) \longrightarrow P(Ob)$ as follows:

$$\mathrm{Int}_B(X) = \{x \in Ob : N_a(x,r) \subseteq X \; \forall a \in B\}$$

$$\mathrm{Cl}_B(X) = \{x \in Ob : N_a(x,r) \cap X \neq \phi \; \forall a \in B\}$$

The classes $\{\mathrm{Int}_B(X) : X \subseteq Ob, B \subseteq At\}$, $\{\mathrm{Cl}_B(X) : X \subseteq Ob, B \subseteq At\}$ and $\{N_B(x,r) : x \in Ob, B \subseteq At\}$ are subbases of a topological spaces denoted τ_I, τ_C and τ_N respectively.

Now let $At = \{a_1, a_2, ..., a_n\}$ and let $\tau_{I_{a_1}}, \tau_{I_{a_2}}, ..., \tau_{I_{a_n}}$, $\tau_{C_{a_1}}, \tau_{C_{a_2}}, ..., \tau_{C_{a_n}}$ and $\tau_{N_{a_1}}, \tau_{N_{a_2}}, ..., \tau_{N_{a_n}}$ be the topologies induced by the subbases $\{\mathrm{Int}_{a_1}(X) : X \subseteq Ob\}$, $\{\mathrm{Int}_{a_2}(X) : X \subseteq Ob\}$, ..., $\{\mathrm{Int}_{a_n}(X) : X \subseteq Ob\}$, $\{\mathrm{Cl}_{a_1}(X) : X \subseteq Ob\}$, $\{\mathrm{Cl}_{a_2}(X) : X \subseteq Ob\}$, ... , $\{\mathrm{Cl}_{a_n}(X) : X \subseteq Ob\}$ and $\{N_{a_1}(x,r) : x \in Ob\}$, $\{N_{a_1}(x,r) : x \in Ob\}$, ... , $\{N_{a_1}(x,r) : x \in Ob\}$, respectively. These topologies called *interior*, *closure* and *neighborhood* topologies respectively.

One of the two attributes a_i, a_j, $i \neq j$ is called *interior-dispensable* in At if, $\tau_{I_{a_i}} = \tau_{I_{a_j}}$, otherwise, a_i or a_j is indispensable in At. Let $\tau_{1,2}, \tau_{2,3}, ..., \tau_{n-1,n}$ be the topologies induced by $\tau_{I_{a_1}} \cup \tau_{I_{a_2}}$, $\tau_{I_{a_2}} \cup \tau_{I_{a_3}}$, ..., $\tau_{I_{a_{n-1}}} \cup \tau_{I_{a_n}}$ if interior topologies are used (the same terminology used if closure topologies or neighborhood topologies is replaced).

Now if $\tau_{I_{At}}$ is the topology induced by $\{\mathrm{Int}_{At}(X) : X \subseteq Ob\}$($\tau_{C_{At}}$ or $\tau_{N_{At}}$ can be used alternately), then when $\tau_{i,j} = \tau_{I_{At}}$ the set $\{a_i, a_j\}$ is a *second order reduct* of At in S. On the other hand, if $\tau_{i,j} \neq \tau_{I_{At}}$ for all $i, j = 1, 2, .., n$ we must calculate the highest topologies $\tau_{1,2,3}$, ..., $\tau_{n-2,n-1,n}$ and the subset $\{a_i, a_j, a_k\}$ is a *third order reduct* of At in S when $\tau_{i,j,k} = \tau_{I_{At}}$. By the same manner, we can define a highly order reducts of At in S.

In each case, the *topological core* of At in S is the intersection of all reducts (intersection of all the same order reducts). This core called the *interior core* and denoted $\mathrm{Core}_{\mathrm{Int}}(At)$. By the same terminology, we can define the *closure core* ($\mathrm{Core}_{\mathrm{Cl}}(At)$) and the *neighborhood core* ($\mathrm{Core}_N(At)$).

Example 3.1. Consider the information system given by Table 3.1 and if we choose $r = 2$, then $N_{a_i}(x,r) = \{y \in Ob :\mid f_{a_i}(x) - f_{a_i}(y) \mid \leq 2\}$, hence we have the following subbases: $\zeta_1 = \{\{x_1, x_2, x_3\}, \{x_1, x_2, x_3, x_4\}, \{x_2, x_3, x_4, x_5\}, \{x_4, x_5\}\}$, $\zeta_2 = \{\{x_1, x_2, x_4\}, \{x_3, x_5\}\}$, $\zeta_3 = \{\{x_1\}, \{x_3, x_4, x_5\}, \{x_2, x_5\}, \{x_3, x_4\}, \{x_2, x_3, x_5\}\}$ and $\zeta_4 = \{\{x_2, x_3, x_4, x_5\}, Ob\}$.

The corresponding bases are:

Ob	a_1	a_2	a_3	a_4
x_1	1	2	9	6
x_2	3	2	6	2
x_3	3	6	3	3
x_4	4	2	2	3
x_5	6	6	5	4

Table 3.1:

$\beta_1 = \{\{x_1, x_2, x_3\}, \{x_1, x_2, x_3, x_4\}, \{x_2, x_3, x_4, x_5\}, \{x_4, x_5\}, \{x_4\}, \{x_2, x_3\}, \{x_2, x_3, x_4\}\}$,

$\beta_2 = \{\{x_1, x_2, x_4\}, \{x_3, x_5\}\}$,

$\beta_3 = \{\{x_1\}, \{x_3, x_4, x_5\}, \{x_2, x_5\}, \{x_3, x_4\}, \{x_2, x_3, x_5\}, \{x_5\}, \{x_3\}, \{x_3, x_5\}\}$, and

$\beta_4 = \{\{x_2, x_3, x_4, x_5\}, \{x_1, x_5\}, \{x_5\}, Ob\}$.

The corresponding topologies are:

$\tau_1 = \{Ob, \phi, \{x_1, x_2, x_3\}, \{x_1, x_2, x_3, x_4\}, \{x_2, x_3, x_4, x_5\}, \{x_4, x_5\}, \{x_4\}, \{x_2, x_3\}, \{x_2, x_3, x_4\}\}$,

$\tau_2 = \{Ob, \phi, \{x_1, x_2, x_4\}, \{x_3, x_5\}\}$,

$\tau_3 = \{Ob, \phi, \{x_1\}, \{x_3, x_4, x_5\}, \{x_2, x_5\}, \{x_3, x_4\}, \{x_2, x_3, x_5\}, \{x_5\}, \{x_3\}, \{x_3, x_5\}, \{x_1, x_2, x_5\}, \{x_1, x_3, x_4, x_5\}, \{x_1, x_2, x_3, x_5\}, \{x_1, x_3, x_4\}, \{x_1, x_5\}, \{x_1, x_3, x_5\}, \{x_1, x_3\}, \{x_2, x_3, x_4, x_5\}\}$ and

$\tau_4 = \{Ob, \phi, \{x_2, x_3, x_4, x_5\}, \{x_1, x_5\}, \{x_5\}\}$.

If we considered the set of all attributes then $\tau_{N_{At}}$ is the discrete topology, but the second order topologies are given such that: $\tau_{1,2} \neq \tau_{N_{At}}$, $\tau_{1,3} = \tau_{N_{At}}$, $\tau_{1,4} \neq \tau_{N_{At}}$, $\tau_{2,3} = \tau_{N_{At}}$, $\tau_{2,4} \neq \tau_{N_{At}}$ and $\tau_{3,4} \neq \tau_{N_{At}}$. Then $\{a_1, a_2\}$ and $\{a_2, a_3\}$ are second order reducts of At and the second order core is given by $\text{Core}_N(At) = \{a_3\}$.

4. Quick Approach for Data Reduction

In [13] when we use Pawlak method for data reduction the following information system (see Table 4.1) has two reducts appears in Table 4.2 and Table 4.3.

U	Headache	Muscle pain	Temperature	Flu
U_1	Yes	Yes	Normal	No
U_2	Yes	Yes	high	Yes
U_3	Yes	Yes	Very high	Yes
U_4	No	Yes	Normal	No
U_5	No	No	high	No
U_6	No	Yes	Very high	Yes

Table 4.1.

U	Muscle pain	Temperature	Flu
U_1, U_4	Yes	Normal	No
U_2	Yes	high	Yes
U_3, U_6	Yes	Very high	Yes
U_5	No	high	No

Table 4.2.

U	Headache	Temperature	Flu
U_1	Yes	Normal	No
U_2	Yes	high	Yes
U_3	Yes	Very high	Yes
U_4	No	Normal	No
U_5	No	high	No
U_6	No	Very high	Yes

Table 4.3.

Formally, the two reducts of Table 4.1 are:

Reduct 1 = {Muscle pain, Temperature}, Reduct 2 = {Headache, Temperature} and have the core, Core = {Temperature}.

Our method for calculating the reducts and the core is shortly suggested by three steps are:

Step 1. Calculate the cardinality of each attribute $a \in At$ denoted by $\mid a \mid$.

Step 2. Determine $\max(\mid a \mid) \forall a \in At$. There are two cases here are:

Case I. Find $\max(\mid a \mid) \forall a \in At$, then the attribute of the maximum cardinality is the core.

Case II. If there exist more than one maximum attribute, then we test the deviation factor of these attributes and the set of attributes of the highest deviation factor is the core.

The deviation factor of an attribute a is a measure of how different values of that attribute take a different decision values. This factor is denoted by η and defined as: $\eta(B) = \mid \{d \in D : \nu_1 = \nu_2 \in a \text{ and } d(\nu_1) \neq d(\nu_2), \forall a \in B\} \mid$, $B \subseteq C$, where D is the decision attribute and C is the condition attributes.

Step 3. Add the core to each subset of the set of all condition attributes after removing from them the core. We take the subsets of the lowest deviation factor as a reducts.

Example 4.1. Consider the same information system given in Table 4.1, then we have: | Muscle pain | = 2, | Headache | = 2 | Temperature | = 3, then we take Temperature as a core and the residue set of the condition attributes is { Muscle pain, Headache }. Now according to Step 3 we will add Temp. to Muscle pain to obtain { Muscle pain, Temperature } and add it to Headache to obtain { Headache, Temperature } and we find that: $\eta(\{\text{Muscle pain}, \text{Temperature}\}) = 0$, and $\eta(\{\text{Headache}, \text{Temperature}\}) = 0$. Hence these two subsets { Muscle pain, Temperature } and { Headache, Temperature } are the reducts.

Example 4.2. Consider the following information system (see Table 4.4),

U	a	b	c	D
u_1	a_0	b_1	c_1	y
u_2	a_1	b_1	c_0	n
u_3	a_0	b_2	c_1	n
u_4	a_1	b_1	c_1	y

Table 4.4.

where $C = \{a, b, c\}$ is the condition attributes and D is the decision attribute. Then we have: $\mid a \mid = 2$, $\mid b \mid = 2$ and $\mid c \mid = 2$ which give no core, but $\eta(a) = 2$, $\eta(b) = 2$ and $\eta(c) = 2$, then $\{b, c\}$ is the unique reduct.

5. Reduction of Data by Covering

Let $R \subseteq Ob \times Ob$ be a binary relation on the universe of an information system S, this relation defines a covering of Ob if it reflexive relation at least. When it is not reflexive, we can make a covering of Ob by adding the negative set defined by this relation.

Let $C = \{C_R(x) : C_R(x) \subseteq Ob\}$ be the covering of Ob by R when R is reflexive, and $C^n = \{C^n_R(x), C_R(x)\}$, where $C_R(x) = R(x) = \{y : xRy\}$ and $C^n_R(x) = Ob - \bigcup_{x \in Ob} C_R(x)$ when R is not reflexive one.

We recall again that any collection of data specified as a structure $(Ob, At, \{V_a : aAt\}, f_a)$ such that Ob is a nonempty set of objects, At is a nonempty set of attributes, $V = \bigcup_{x \in Ob} V_a$ is a nonempty set of values and f_a is a function of Ob into $2^{V_a \setminus \{\phi\}}$, is referred to as a multi-valued information system.

In this section we assume that with every attribute $a \in At$ is related a reflexive relation R_a. For simplicity, this relation shall be defined in the following way. Let $a \in At$ and $B \subseteq At$ then xR_ay iff $f_a(x) \cap f_a(y) \neq \phi$, and xR_By iff xR_ay $\forall a \in B$. Also, the relation R^W_B defined by: xR^W_By iff $\exists a \in B$ xR_ay. The relation R^W_B is reflexive relation and called the weak relation derived by the strong relation R_a. The coverings C_{R_B} and $C_{R^W_B}$ are subbases of two topologies called the strong and weak topologies and, denoted by $\tau_{C_{R_B}}$ and $\tau_{C_{R^W_B}}$ respectively. The class $\{f_a(x) : a \in At\}$ shall be called an information about the object x, or a record of x. We shall say that two records determined by x, y are strongly similar with respect to $\tau_{C_{R^W_B}}$ iff $\forall a \in At$ $f_a(x) \cap f_a(y) \neq \phi$. Also two records $\{f_a(x) : a \in B \subseteq At\}$ and $\{f_b(x) : b \in B \subseteq At\}$ are $\tau_{C_{R_B}}$ strongly similar with respect to the set $B \subseteq At$ iff $\forall a \in B$ $f_b(x) \cap f_b(y) \neq \phi$. Two objects x, y are weakly $\tau_{C_{R^W_B}}$ similar if for some $a \in B$ $f_a(x) \cap f_a(y) \neq \phi$.

The set of attributes $Y \subseteq At$ depends on the set $X \subseteq At$ with respect to the strong similar topology $\tau_{C_{R^W_B}}$ if and only if $C_{R_X} \leq C_{R_Y}$. In the same way we can define dependency of attributes with respect to the weak similar topology $\tau_{C_{R^W_B}}$: $X \xrightarrow{R^W_{At}} Y$ iff $C_{R^W_X} \leq C_{R^W_Y}$.

By information system in general sense, we shall understand the family of operators on subsets of a given universe. The operators are usually called lower and upper approximation operators. They can be defined using equivalence relations, partitions, covers, similar relations, topology, etc.

Let C_{R_B} be a covering of the universe Ob of an information system S. For every $X \subset Ob$ we define the n-th upper approximation of X using C_{R_B} by:

$C^{-1}(X) = \bigcup\{C \in C_{R_B} : C \cap X \neq \phi\}$, $C^{-2}(X) = C^{-1}(C^{-1}(X))$, $C^{-3}(X) = C^{-1}(C^{-2}(X))$, ... , $C^{-n}(X) = C^{-1}(C^{-n-1}(X))$. Analogously $C^n(X) = -C^{-n}(-X)$. is called n-th lower approximation by C_{R_B} of the set X.

Let $F : P(At) \longrightarrow P(Ob \times Ob)$ be arbitrary function satisfying the conditions: For every $B \subseteq At$ $F(B) = \bigcap\{F(\{b\}) : b \in B\}$ and $F(\phi) = Ob \times Ob$.

Now for any subset $X \subseteq At$ we say that X is $F-$independent iff for every set $Y \subset X$ it holds $F(Y) \neq F(X)$. Otherwise, we say that X is $F-$dependent. The set $Y \subset X$ is called $F-$reduct of X iff $F(Y) = F(X)$ and Y is $F-$independent the class of all reducts of any subset $B \subseteq At$ will be denoted by $\text{Red}_{F(B)}$ and the class of all reducts of At will be denoted by Red_F.

Let $B \subseteq At$ and $a \in B$, then a is called $F-$dispensable in B if $F(B) = F(B - \{a\})$. Otherwise a is called $F-$indispensable. The core of B is defined by:

$\text{Core}(B) = \{b \in B : b \text{ is } F - \text{indispensable in } B\}$.

Now, for any $B \subseteq At$, we shall say that the set B is R_B- independent iff for every set $Y \subseteq B$ it holds $C_{R_Y} \neq C_{R_B}$. Otherwise we say that B is dependent with respect to R_B. The subset $Y \subseteq B$ is a reduct of B with respect to $\tau_{C_{R_B}}$ iff $C_{R_Y} = C_{R_B}$ and Y is R_B-independent. By R_B-Core of B, we mean the set $\text{Core}_{R_B}(B) = \{b \in B : be \text{ is } R_B - \text{indispensable in } B\}$.

Fact 1. For every $B \subseteq At$,

$$\text{Core}_{R_B}(B) = \bigcap\{Q : Q \text{ is a reduct of } B \text{ with respect to } R_B\}.$$

Example 5.1. Consider the information system given by Table 5.1.

Here $Ob = \{P_1, P_2, ..., P_8\}$ and $At = \{M_1, M_2, ..., M_5\}$. Then:

$C_{R_{M_1}} = \{\{P_1, P_4, P_5, P_7, P_8\}, \{P_2, P_3, P_4, P_5, P_6, P_8\}, \{P_2, P_2, P_5, P_8\},$
$\{P_1, P_2, P_3, P_6, P_8\}, \{P_1, P_2, P_3, P_5, P_7, P_8\}, \{P_2, P_4, P_6\}, \{P_1, P_5, P_7\},$
$\{P_1, P_2, P_3, P_4, P_5, P_8\}\}$.t

$C_{R_{M_2}} = \{\{P_1, P_2, P_4, P_5, P_7\}, \{P_1, P_2, P_4, P_5, P_6, P_7, P_8\}, \{P_3, P_6\},$
$\{P_2, P_3, P_4, P_6, P_8\}, \{P_1, P_2, P_4, P_5, P_7, P_8\}, \{P_2, P_5, P_6, P_7, P_8\}\}$.

$C_{R_{M_3}} = \{\{P_1, P_3, P_6\}, \{P_2, P_4, P_8\}, \{P_1, P_3, P_5, P_7\}, \{P_2, P_4, P_6, P_8\},$
$\{P_3, P_5, P_7\}, \{P_1, P_4, P_6\}, \{P_3, P_5, P_7\}, \{P_2, P_4, P_8\}\}$.

$C_{R_{M_4}} = \{\{P_1, P_4\}, \{P_2, P_3, P_4\}, \{P_2, P_3, P_4, P_7, P_8\}, \{P_1, P_2, P_3, P_4\},$
$\{P_5\}, \{P_6, P_8\}, \{P_3, P_7, P_8\}, \{P_3, P_6, P_7, P_8\}\}$.

$C_{R_{M_5}} = \{\{P_1\}, \{P_2, P_4, P_8\}, \{P_3, P_5, P_7\}, \{P_2, P_4, P_6, P_7\}, \{P_3, P_5, P_7\},$
$\{P_4, P_6, P_7, P_8\}, \{P_3, P_4, P_5, P_6, P_7\}, \{P_2, P_6, P_8\}\}$. Let $B = \{M_1, M_2\}$, then
$C_{R_{M_B}} = \{\{P_1, P_4, P_5, P_7\}, \{P_2, P_4, P_6, P_8\}, \{P_3\}, \{P_1, P_2, P_4\},$
$\{P_1, P_2, P_5, P_7, P_8\}, \{P_2, P_6\}, \{P_1, P_5, P_7\}, \{P_2, P_5, P_8\}\}$.

Ob	M_1	M_2	M_3	M_4	M_5
P_1	$\{2,5\}$	$\{b\}$	$\{1,a\}$	$\{2\}$	$\{c\}$
P_2	$\{3,4\}$	$\{b,c\}$	$\{3,c\}$	$\{b\}$	$\{2,3\}$
P_3	$\{3\}$	$\{a\}$	$\{1,2,d\}$	$\{3,b,c\}$	$\{d\}$
P_4	$\{4,5\}$	$\{b\}$	$\{3,b\}$	$\{2,a,b\}$	$\{2,a\}$
P_5	$\{2,3\}$	$\{b,c\}$	$\{2,d\}$	$\{4,d\}$	$\{5,d\}$
P_6	$\{4\}$	$\{a,c\}$	$\{a,b\}$	$\{5\}$	$\{4,a\}$
P_7	$\{2\}$	$\{b,d\}$	$\{d\}$	$\{3,c\}$	$\{a,d\}$
P_8	$\{3,5\}$	$\{c,d\}$	$\{3,4\}$	$\{5,c\}$	$\{3,4\}$

Table 5.1:

Now if we take $F(Y) = R_Y, \forall y \in B$, then for $Y_1 = \{M_1\}$, $F(\{M_1\}) = R_{M_1}$ and for $Y_2 = \{M_2\}$, $F(\{M_1\}) = R_{M_1}$. But $C_{R_{M_1}} \neq C_{R_{M_2}} \neq C_{R_{M_B}}$ then $B = \{M_1, M_2\}$ is $F-$independent set. Let $D = \{M_1, M_3\}$, then

$C_{R_D} = \{\{P_1\}, \{P_2, P_4, P_8\}, \{P_3, P_5\}, \{P_2, P_4, P_6, P_8\}, \{P_3, P_5, P_7\}, \{P_4, P_6\}, \{P_5, P_7\}$.

Also $C_{R_D} \neq C_{R_{M_1}} \neq C_{R_{M_3}}$, then D is $F-$independent.

Let $E = \{M_2, M_3\}$, then

$C_{R_E} = \{\{P_1\}, \{P_2, P_4, P_8\}, \{P_3\}, \{P_2, P_4\}, \{P_5, P_7\}, \{P_6\}, \{P_2, P_8\}\}$.

Let $G = \{M_1, M_2, M_3\}$, then

$C_{R_D} = \{\{P_1\}, \{P_2, P_4, P_8\}, \{P_3\}, \{P_2, P_4\}, \{P_5, P_7\}, \{P_6\}, \{P_2, P_8\}\}$.

Then $F(G) = F(G - \{M_1\}) = F(E)$. Then the attribute M_1 is $F-$dispensable in G. Also, the Core$(G) = \{M_2, M_3\}$.

References

[1] M.E. Abd El-Monsef, *Studies on Some Pretopological Concepts*, Ph.D. Thesis, Tanta University (1980).

[2] D. Andrijevic, On the topology generated by preopen sets, *Math.*, **39** (1987), 367-376.

[3] O. Ahlqvist, J. Keukelaar, K. Oukbir, Rough classification and accuracy assessment, *International Journal of Geographical Information Science*, **14**, No. 5 (2000), 475-496.

[4] E. Clementini, P. Di Felice, Approximate topological relations, *International Journal of Approximate Reasoning*, **16**, No. 2 (1997), 173-204,.

[5] E. Flapan, *When Topology Meets Chemistry*, Cambridge University Press (2000).

[6] J. Jelonek, K. Krawiec, R. Slowinski, Rough set reduction of attributes and their domains for neural networks (2002), 339-347.

[7] S. Marcus, Tolerance rough sets, $\hat{C}$ech topologies, learning processes, *Bull. Polish Acad. Sci. Tech. Sci.*, **42**, No. 3 (1994), 471-487.

[8] I.J. Nagata, *Modern General Topology*, North Holland Pub. Co., Amsterdam (1968).

[9] A. Wiweger, On topological rough sets, *Bull. Pol. Ac. Math.*, **37** (1989), 51-62.

[10] Z. Pawlak, Information systems theoretical foundations, *Information Systems*, **6** (1981), 205 - 218.

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

[12] Z. Pawlak, Reasoning about data – A rough set perspective, In: *Polkowski and Skowron* (1998), 25-34.

[13] Z. Pawlak, W. Marek, Rough sets and information systems, *ICS. PAS. Reports* (1981), 481-485.

[14] Z. Pawlak, Rough sets, algebraic and topological approach, *ICS. PAS. Reports* (1982), 99-104.

[15] Z. Pawlak, On rough relations, *Bull. Pol. Ac. Tec. Sciences*, **34** (1986), 9-10.

[16] Z. Pawlak, Rough sets, rough relations and rough functions, *Bull. Pol. Ac. Math.* (1996), 15-19.