A NEW TOPLOGICAL REDUCATION OF INCOMPLETE INFORMATION SYSTEMS IN MEDICAL DIAGNOSIS

M.Lellis Thivagar[1], Carmel Richard[2]

Abstract: Attribute Reduction is a basic issue in rough set research. This paper is an attempt to deal with attribute reduction in tolerance based consistent incomplete information systems . The concept of tolerance relation and tolerance class are used to introduce topological reduction in terms of base and it is applied in medical diagnosis.

Keywords: Core, Incomplete information system, Rough Sets, Tolerance class, Tolerance relation.

1. INTRODUCTION:

Rough set theory, introduced by Zdzislaw Pawlak, is a new mathematical approach to imperfect knowledge which is a crucial issue for computer scientists.. Several interesting applications of the theory have come up, in particular, in Artificial Intelligence and Cognitive Sciences. This theory deals with the approximation of sets or concepts by means of equivalence relations and is considered as one of the first non-statistical approaches in data analysis. Due to increased existence of incomplete information systems in real life, many authors employed extensions of Pawlak's rough set model to reason in incomplete information systems with missing attribute values. An incomplete information system is an extended model of rough set theory based on intersecting granules, called tolerance granules. To derive certain possible decision rules in incomplete information system, Leung et al [2] developed a new rough set approximation by using labeled blocks. Not all conditional attributes in an information system are necessary to depict the decision attribute before decision rules are generated. Attribute reduction is an outstanding contribution made by rough set research [4]. In this paper we attempt to apply topological reduction of attributes in incomplete information systems to find key symptoms for two common diseases among women, Hyperthyroidism and Osteoporosis using the concept of basis.

2. PRELIMINARIES

Definition 2.1 [6]: An information system is of the form (U, A,{Va }, fa) a$\in$ A where,

- U ia a non-empty finite set of objects, called the universe
- A is a non- empty finite set of attributes
- V_a is the attribute value set of an attribute a $\in$ A
- $f_a : U\rightarrow V_a$ is called the information function.

If f_a (x) is equal to a missing value for some x ∈ U and a ∈ A, then the information system is called an incomplete information system (IIS). Otherwise it is a complete information system (CIS). A missing value is denoted by (*). That is, an IIS is of the form (U, A, {Va}, fa) a ∈A where * ∈ ∪ Va. For any B ⊆ A, there is a binary relation denoted by TR(B) and given by TR(B) = { (x,y) ∈ U x U : f_a (x) = f_a(y) or f_a (x)=* or f_a (y) = * for any a∈ B}. Then TR(B) is a tolerance relation(reflexive and symmetric) on B . S_B (x) denotes the maximal set of objects which are possibly indiscernible with x by this tolerance relation. That is, S_B(x) = {y∈U: (x,y) ∈ TR(B)} if x∈ U and B ⊆ A.

Definition 2.2[6]: Let U be a non-empty finite universe and X be a subset of U. Given a tolerance relation on a subset of B of the attribute set A, the *lower approximation* is given by, L_B(X) = {x∈U:S_B(x) ⊆ X } and the *upper approximation* of X is given by U_B(X) = {x∈U : S_B (x) ∩ X ≠ φ}. That is,the objects in the lower approximation of X certainly belong to X whereas the objects in U_B (X) possibly belong to X . A set Y is said to be *rough*, if Y is of the form (A_1, A_2) where A_1 = L_B(X) and A_2 = U_B(X) for some subset X of U.

Definition 2.3 [6]: If IIS= (U,A) with B ⊆ A and X ⊆ U, then X is said to be a tolerance class with respect to B if (x, y) ∈ TR(B) for any x, y ∈ X.

Definition 2.4 [6]: U/ TR (B) denotes the set of all maximal tolerance classes with respect to B and is called as a full cover of U.

Remark 2.5: A given tolerance relation may give rise to many covers of an universe, but only one full cover. Thus, there exists a one-to-one correspondence between tolerance relations and full covers.

Example 2.6:

Consider the following information table where U={x_1,x_2,…x_5 } and A= {$a_=$,a_2,a_3}

U	a_1	a_2	a_3
x_1	2	*	3
X_2	3	2	1
x_3	2	1	*
x_4	*	*	3
x_5	1	1	2

We note that f_{a_2}(x_1), f_{a_3}(x_3), f_{a_1}(x_4) and f_{a_2}(x_4) are missing attribute values. Therefore, the given information system is an IIS.

Also S_A(x_1) = {x_1,x_3, x_4} = S_A(x_3) = S_A(x_4) ; S_A(x_2) = {x_2} ; S_A(x_5) = {x_5}

Let X= { x_1,x_2,x_4}. Then, L_A(X)= {x_2} ; U_A(X) = { x_1,x_2,x_3,x_4}.

U/ TR (a_1) = { {x_1,x_3,x_4}, {x_2,x_4}, {x_4,x_5} }, U/TR(a_2) = { {x_1,x_2, x_4}, {x_1,x_3,x_4,x_5}}, U/TR(a_3)= {{ x_2,x_3},{x_3,x_5},{x_1,x_3,x_4}.

Definition 2.7 [6]: If IIS= (U, A), a subset B of A is a core of IIS if U/TR (B) = U/TR (A) and for every B'$\subset$ B, U/TR(B') $\neq$ U/TR(A).

3. ATTRIBUTE REDUCTION IN IIS

Definition 3.1 : If U is a non-empty set of objects and R is a tolerance relation on U, a sub-knowledge base for each r in R is given by $S_r = \{ S_C(x) : x \in U\}$

Definition 3.2 : A sub-base for the topology τ on U is given by $S_R = \cup_{r \in R} S_r$ and a base for is given by $\beta_R = \cap_{S_x \in S_R} S_x$ (with maximality of each S_x taken into consideration).

Remark 3.3 : Let U and A be two non-empty finite sets, where U is the universe and A, the set of attributes. With every attribute a in A, we associate a set Va of its values. An interesting question is whether there is a subset of attributes in the information system which, by itself, can fully characterize the knowledge in the database. An IIS may possess more than one reduct. However, in real life situations, we are interested in a particular reduct, such as the minimal reduct or any reduct containing key attributes. Such an attribute set is called the *core* , which is introduced in terms of base as follows.

Definition 3.4 : Let (U,A) be an IIS, where A is divided into a set C of conditional attributes and a set D of decision attribute. A subset R of C is said to be a *core,* if $\beta_R = \beta_C$ and $\beta_R \neq \beta_{R-\{r\}}$ for all $r \in R$. That is, a core of R is a minimal subset of attributes which is such that none of its elements can be removed without affecting the classification power of attributes.

4. TOPOLOGICAL REDUCTION OF ATTRIBUTES IN HYPERTHYROIDISM AMONG WOMEN

Thyroid hormones stimulate the metabolism of cells. They are produced by the thyroid gland. The thyroid gland is located in the lower part of the neck, below the Adam's apple. The gland wraps around the windpipe (trachea) and has a shape that is similar to a butterfly formed by two wings (lobes) and attached by a middle part (isthmus). Hyperthyroidism is a condition in which an overactive thyroid gland produces an excessive amount of thyroid hormones that circulate in the blood. Thyrotoxicosis is a toxic condition that is caused by an excess of thyroid hormones from any cause.

Thyrotoxicosis can be caused by an excessive intake of thyroid hormone or by overproduction of thyroid hormones by the thyroid gland. Common symptoms include high blood pressure, weight loss, goiter and irregular menstrual flow. Consider the following information table giving information about six patients

ISBN - 978-93-81583-56-2

Table I

Patients	BP (B)	Weight (W)	Goiter(G)	Menstrual Flow(M)	Hyperthyroid (HT)
P_1	Low	Normal	Enlarged	Increased	No
P_2	*	Low	Normal	Increased	No
P_3	Normal	Heavy	Enlarged	Decreased	No
P_4	High	Heavy	Enlarged	Normal	Yes
P_5	*	Low	Normal	*	No
P_6	High	Heavy	Enlarged	*	Yes

Information Table of patients with / without Hyperthyroid

The given information system is incomplete and is given by (U,A) where U = $\{P_1,P_2,P_3,P_4,P_5,P_6\}$ and A ={BP(B), Weight(W),Goiter (G), Menstrual Flow(M), Hyperthyroid(HT)}. The attribute set A is divided into two subsets –C,of conditional attributes {BP, Weight, Goiter, Menstrual Flow} and D,of decision attribute {Hyperthyroid}.

U/TR(C) = $\{ \{ P_1\}, \{ P_2, P_5 \},\{ P_3 \}, \{P_4,P_6\} \}$ and U/TR (D) = $\{\{ P_1, P_2, P_3, P_5\}, \{P_4,P_6\}\}$. The sub- knowledge base for each $a\in A$ is given by U/TR(a) $=\{S_B(x) : x \in U\}$ where B={a}.

Therefore, the sub-knowledge base for each of the condition attributes is given by

U/TR(B) = $\{ \{P_1,P_2,P_5\},\{P_2,P_3,P_5\},\{P_2,P_4,P_5,P_6\} \}$

U/TR (W) = $\{ \{ P_1\}, \{P_2,P_5\}, \{P_3,P_4,P_6\} \}$

U/TR (G) = $\{ \{P_1,P_3,P_4,P_6\}, \{P_2,P_5\} \}$

U/TR (M) = $\{ \{P_1,P_2,P_5,P_6\}, \{P_3,P_5,P_6\},\{ P_4,P_5,P_6\} \}$

Therefore the sub- base for the topology of conditional attributes is given by $S_c = \{\{P_1,P_2,P_5\}, \{P_2,P_3,P_5\}, \{P_2,P_4,P_5,P_6\}, \{P_1\}, \{P_2,P_5\}, \{P_3,P_4,P_6\}, \{P_1,P_3,P_4,P_6\}, \{P_1,P_2,P_5,P_6\}, \{P_3,P_5,P_6\}, \{ P_4, P_5,P_6\}\}$. And the base for the topology is given by $\beta_C =\{ \{P_1\}, \{P_2,P_5\},\{P_3\},\{ P_4,P_6\}\}$ (=U/TR(C)).Since not all the symptoms are necessary to decide on the occurrence of hyperthyroidism, we remove the conditional attributes one by one from C and find the core which is the minimal set of attributes corresponding to the key symptoms for the disease.

$\beta_{C-\{B\}} = \beta\{W,G,M\}= \{\{P_1\},\{P_2,P_5\}, \{P_3,P_6\}, \{P_4,P_6\}\} \neq \beta_C$ and

$\beta_{C-\{W\}} = \beta_{\{B,G,M\}} = \{\{ P_1\},\{P_2,P_5\},\{P_3\}, \{P_4,P_6\}\} = \beta_{C.}$

Let R = C- {w}. then, $\beta_{R-\{B\}} = \beta_{\{G,M\}}=\{\{P_1,P_6\}, \{P_2,P_5\},\{P_3,P_6\}, \{P_4,P_6\}\} \neq \beta_C$, $\beta_{R-\{G\}} = \beta_{\{B,M\}} = \{\{ P_1,P_2,P_5\}, \{P_3,P_5\},\{P_4,P_5,P_6\}\} \neq \beta_C$, $\beta_{R-\{M\}} = \beta_{\{B,G\}} = \{\{P_1\}, \{P_2,P_5\},\{P_3\},\{P_4,P_6\}\} = \beta_C$

Therefore *core* = R- {M} = {B,G}

Observation: 'BP' and 'GOITER ' are the key symptoms that have close connection to the disease 'hyperthyroid' that draw much more attention than the other ones.

5. ATTRIBUTE REDUCTION IN OSTEOPOROSIS

Osteoporosis means "porous bones." For patients with osteoporosis, the bones don't look any different, but they lose substance as well as calcium and other minerals. As a result, the bones become weak and are more likely to fracture, particularly if you fall. Osteoporosis develops slowly over several years. It is likely there will be no warning signs or symptoms until a minor fall or sudden impact causes a bone fracture. A simple cough or a sneeze may cause the fracture of a rib or the partial collapse of one of the bones of the spine. One visible sign of osteoporosis is the characteristic stooping (bent forward) position that develops in older people. It happens when the bones in the spine are fractured, making it difficult to support the weight of the body. The appearance of a widow's hump or a fractured wrist or hip from a fall may be the first actual symptoms of osteoporosis. When the bones lack sufficient strength to hold the weight of your body, we experience joint / back ache . Unexplained bone or joint pain may raise the consideration of a bone health problem. Some patients experience cramps in the legs at night.

Table2: Table of patients with / without Osteoporosis

Patients	Joint or back ache (A)	Stooped Figure (S)	Low energy fractures (F)	Cramps in legs (CL)	Osteo-porosis
P_1	Severe	Yes	Yes	No	Yes
P_2	No	Yes	No	*	No
P_3	Mild	Yes	No	Yes	No
P_4	*	Yes	Yes	*	Yes
P_5	Severe	Yes	Yes	*	Yes
P_6	*	No	No	Yes	No

The given information system is incomplete and is given by (U,A) where

U= $\{P_1, P_2, P_3,P_4,P_5,P_6\}$ and A ={Joint or back ache(A), Stooped figure (S) , Low energy fractures (F),Cramps in legs(CL), Osteoporosis}, which is divided into two subsets, namely a set C, of conditional attributes {A,S,F,CL} and D,of decision attribute {Osteoporosis}. U/TR(C) = $\{\{P_1, P_4, P_5\},\{P_2\}, \{P_3\},\{P_6\}\}$ and U/TR (D) = $\{\{P_1,P_4,P_5\},\{P_2,P_3,P_6\}\}$.Then, POS_C (D) = $\{P_1,P_2,P_3,P_4,P_5,P_6\}$ = U and hence the IIS is consistent. Therefore, the sub-knowledge base for each of the condition attributes is given by

ISBN - 978-93-81583-56-2

$U/TR(A) = \{\{P_1, P_4, P_5, P_6\}, \{P_2, P_4, P_6\}, \{P_3, P_4, P_6\}\}$

$U/TR(S) = \{\{P_1, P_2, P_3, P_4, P_5\}, \{P_6\}\}$

$U/TR(F) = \{\{P_1, P_4, P_5\}, \{P_2, P_3, P_6\}\}$

$U/TR(CL) = \{\{P_1, P_2, P_4, P_5\}, \{P_2, P_3, P_4, P_5, P_6\}\}$

Therefore the sub- base for the topology of conditional attributes is given by $Sc = \{\{P_1, P_4, P_5, P_6\}, \{P_2, P_4, P_6\}, \{P_3, P_4, P_6\}, \{P_1, P_2, P_3, P_4, P_5\}, \{P_6\}, \{P_1, P_4, P_5\}, \{P_2, P_3, P_6\}, \{P_1, P_2, P_4, P_5\}, \{P_2, P_3, P_4, P_5, P_6\}\}$.

And the base for the topology on U is given by

$\beta_C = \{\{P_1, P_4, P_5\}, \{P_2\}, \{P_3\}, \{P_6\}\}$.

$\beta_{C-\{A\}} = \{\{P_1, P_4, P_5\}, \{P_2, P_3\}, \{P_6\}\} \neq \beta_C$.

$\beta_{C-\{S\}} = \{\{P_1, P_4, P_5\}, \{P_2\}, \{P_3, P_6\}\} \neq \beta_C$.

$\beta_{C-\{F\}} = \{\{P_1, P_4, P_5\}, \{P_2, P_4\}, \{P_3, P_4\}, \{P_6\}\} \neq \beta_C$.

$\beta_{C-\{CL\}} = \{\{P_1, P_4, P_5\}, \{P_2\}, \{P_3\}, \{P_6\}\} = \beta_C$.

Let $R = C - \{CL\} = \{A,S,F\}$. Then, $\beta_R = \beta_C$ and $\beta_{R-\{A\}} = \{\{P_1, P_4, P_5\}, \{P_2, P_3\}, \{P_6\}\} \neq \beta_C$, $\beta_{R-\{S\}} = \{\{P_1, P_4, P_5\}, \{P_2, P_6\}, \{P_3, P_6\}\} \neq \beta_C$ and $\beta_{R-\{F\}} = \{\{P_1, P_4, P_5\}, \{P_2, P_4\}, \{P_3, P_4\}, \{P_6\}\} \neq \beta_C$. Therefore, *core* = { A,S,F}

Observation : From the definition of *core* , **'Joint or back ache'**, **' Stooped figure'** and **' Low energy fractures'** are the key symptoms that have close connection to the disease under consideration.

6. CONCLUSION

In this work, the concept of basis has been applied to find the deciding factors of two common diseases among women 'Hyperthyroid' and 'Osteoporosis'. We could find that '**BP**' and '**Goiter'** are the key symptoms for hyperthyroid and **'Joint or back ache', 'Stooped figure'** and **'Low energy fractures'** are the deciding factors for Osteoporosis. It is also seen that from a clinical point of view, this model is on par with the medical experts with respect to the diseases analyzed here. The proposed topological reduction can be applied to more general and complex information systems for future research. The rough set model is based on the original data only and does not need any external information, unlike probability in statistics or grade of membership in the fuzzy set theory. It is also a tool suitable for analyzing not only quantitative attributes but also qualitative ones. The results of the rough set model are easy to understand, while the results from other methods need an interpretation of the technical parameters. Thus it is advantageous to use rough set model in real life situations.

7. REFERENCES

[1] E. F. Lashin, T. Medhat, “Topological reduction of information systems”, Chaos, Solitons and Fractals, Vol. 25, 2005, pp 277-286.

[2] Y. Leung, W.-Z. Wu, W.-X. Zhang,” Knowledge acquisition in incomplete information systems: a rough set approach”, European Journal of Operational Research , Vol.168 2006 , pp164–180.

[3] Wei-Zhi Wu, Mei Zhang, Huai-Zu Li and Ju-Sheng Mi , “ Knowledge reduction in random information systems via Dempster–Shafer theory of evidence”, Information Sciences, Vol. 174, 2005, pp143–164

[4] Z. Pawlak, “Rough Sets: Theoretical Aspects of Reasoning about Data” , Kluwer Academic Publishers, Boston, 1991.

[5] Zdzislaw Pawlak ,“Rough set theory and its applications”, Journal of Telecommunications and Information Technology”, Vol.3, 2002 , pp. 7-10.

[6]. Zuqiang Meng , Zhongzhi Shi , “A fast approach to attribute reduction in incomplete decision systems with tolerance relation-based rough sets”, Information Sciences Vol. 179 2009, pp 2774-2793

* * * *

[1]*Author1 : School of Mathematics, Madurai Kamaraj University, Madurai-625021- India.*
mlthivagar@yahoo.co.in
[2]*Author 2 : Department of Mathematics, Lady Doak College, Madurai - 625002-India.*
carmel09richard@gmail.com

ISBN - 978-93-81583-56-2