ADV MATH SCI JOURNAL

Advances in Mathematics: Scientific Journal **9** (2020), no.2, 533–538
ISSN: 1857-8365 (printed); 1857-8438 (electronic)
Special Issue on ICCSPAM-2020; https://doi.org/10.37418/amsj.9.2.1

ANALYSING DIABETES USING ROUGH SETS

S. ARJUN RAJ[1] AND M. VIGNESHWARAN

ABSTRACT. In this article we use the rough set theory to generate the set of decision concepts in order to solve a medical problem.Based on officially published data by International Diabetes Federation (IDF), rough sets have been used to diagnose Diabetes.The lower and upper approximations of decision concepts and their boundary regions have been formulated here.

1. INTRODUCTION

In mathematics, topology [1] is concerned with the properties of space that are preserved under continuous deformations, such as stretching, crumpling and bending, but not tearing or gluing. Rough sets was introduced by Zdzislaw Pawlak [2] in 1982. It is a formal theory derived from fundamental research on logical properties of information systems. Rough set theory has been a methodology of database mining or knowledge discovery in relational databases. In its abstract form, it is a new area of uncertainty in mathematics closely related to fuzzy theory.The main goal of the rough set [2] analysis is the induction of (learning) approximations of concepts. It offers mathematical tools to discover patterns hidden in data.

[1] *corresponding author*

Key words and phrases. Rough sets, Approximations, Concept, Information table, Patient data, Analysis.

2. Preliminaries

Definition 2.1. *[1]*

A topology [1] on a set X is a collection τ of subsets of X having following properties:

- *ϕ and $X \in \tau$*
- *The union of the elements of any sub-collection of τ is in τ*
- *The intersection of the elements of any finite sub-collection of τ is in τ*

Definition 2.2. ***Approximations*** *[3]*

It is defined by using a system of definable sets, a formal approximation of a crisp set defined by its two approximations namely: Upper approximation and Lower approximation.

Upper approximation

It is the set of objects which possibly belong to the target set.

$$\overline{R}X = \bigcup\{Y \in U/R : Y \bigcap X \neq \phi\}$$

Lower approximation

It is the set of objects that positively belong to the target set.

$$\underline{R}X = \bigcup\{Y \in U/R : Y \subseteq X\}.$$

Definition 2.3. ***Boundary Region*** *[2]*

A set is said to be rough if its boundary region is non-empty, otherwise the set is crisp.

$$BN_R(X) = \overline{R}(X) - \underline{R}(X)$$

A rough set is composed of two crisp sets, one representing a lower boundary of the target set and the other representing an upper boundary of the target set.

Definition 2.4. ***Indiscernibility*** *[4]*

Tables may contain many objects having the same features. A way of reducing table size is to store only one representative object for every set of objects with same features.These objects are called indiscernible objects or tuples. With any P subset A there is an associated equivalence relation IND(P):

$$\textit{IND(P)} \equiv \{(x, y) \in \bigcup^{2} |\forall a \in P, a\,(x) = a\,(y)\},$$

where IND(P) is called indscernibility of relation. Here x and y are indiscernible from each other by attribute P.

Definition 2.5. ***Interior*** *[4]*

In topology, the interior of a subset of a topological space is the union of all open subsets of that set.

Definition 2.6. ***Closure*** *[3]*

The closure of a subset S of points in a topological space consists of all points in S together with all limit points of S.

Definition 2.7. ***Diabetes***

Diabetes is a disease that affects a bodies ability to produce or use insulin. Insulin is a hormone that helps to control body glucose levels. When a body turns the food we eat into energy (also called sugar or glucose), insulin is released to help transport this energy to the cells. Insulin acts as a key. If a body produce little or no insulin, or are insulin resistant, too much sugar remains in the blood. Blood glucose levels are higher than normal for individuals with diabetes.

3. Identifying diabetes using Rough set

In this section we have applied the Rough Set theory to analyze diabetes using topological reduction of attributes in information system.

Algorithm

The following algorithm is developed to find the deciding factors or core to pick the minimum number of attributes necessary for the classification of objects.

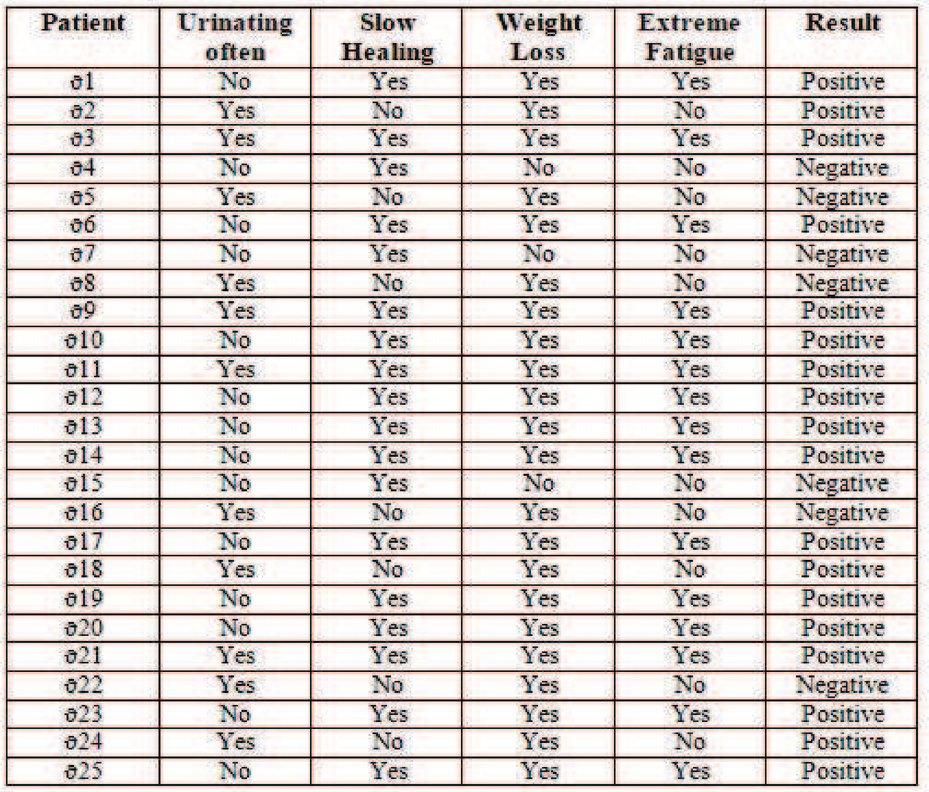
Step 1: Given a finite universal set U, a finite set X of attributes , an equivalence relation R on U corresponding to a subset X of U which represent the data as an information table, columns of which are labeled by attributes and rows by objects and entries of the table are attribute values.

Step 2: Find the lower approximation,of the set X with respect to R.

Step 3: Find the upper approximation, of the set X with respect to R.

Step 4: Find the boundary region of the set X with respect to R,which can be classified neither as X nor not in X with respect to R .

Consider the table of data about the patients with the following symptoms. Urinating Often, Slow Healing, Weight Loss and Extreme Fatigue.



Patient	Urinating often	Slow Healing	Weight Loss	Extreme Fatigue	Result
$\vartheta 1$	No	Yes	Yes	Yes	Positive
$\vartheta 2$	Yes	No	Yes	No	Positive
$\vartheta 3$	Yes	Yes	Yes	Yes	Positive
$\vartheta 4$	No	Yes	No	No	Negative
$\vartheta 5$	Yes	No	Yes	No	Negative
$\vartheta 6$	No	Yes	Yes	Yes	Positive
$\vartheta 7$	No	Yes	No	No	Negative
$\vartheta 8$	Yes	No	Yes	No	Negative
$\vartheta 9$	Yes	Yes	Yes	Yes	Positive
$\vartheta 10$	No	Yes	Yes	Yes	Positive
$\vartheta 11$	Yes	Yes	Yes	Yes	Positive
$\vartheta 12$	No	Yes	Yes	Yes	Positive
$\vartheta 13$	No	Yes	Yes	Yes	Positive
$\vartheta 14$	No	Yes	Yes	Yes	Positive
$\vartheta 15$	No	Yes	No	No	Negative
$\vartheta 16$	Yes	No	Yes	No	Negative
$\vartheta 17$	No	Yes	Yes	Yes	Positive
$\vartheta 18$	Yes	No	Yes	No	Positive
$\vartheta 19$	No	Yes	Yes	Yes	Positive
$\vartheta 20$	No	Yes	Yes	Yes	Positive
$\vartheta 21$	Yes	Yes	Yes	Yes	Positive
$\vartheta 22$	Yes	No	Yes	No	Negative
$\vartheta 23$	No	Yes	Yes	Yes	Positive
$\vartheta 24$	Yes	No	Yes	No	Positive
$\vartheta 25$	No	Yes	Yes	Yes	Positive

Table 1:Set of Data gathered from IDF
(International Diabetes Federation).

In the above table containing the information of patients suffering from the symptoms; Urinating Often, Slow Healing, Weight Loss and Extreme Fatigue.The above table is known as information systems, attribute values tables or information tables here we use the term information table.

Here we are given with the data of twenty-five patients with same symptoms. As shown in table 1.The columns of the table are labeled by attributes (symptoms) and rows objects (patients), where as entries of the table are attribute values.Thus each row of the table can be seen as information about specific patient. For example, patient $\vartheta 2$ is characterized in the table by the following attribute value set [Urinating Often - Yes], [Slow Healing-no], [Weight Loss - yes], [Extreme Fatigue - No] which form the information about the patient.

In the above table patients $\vartheta 2$, $\vartheta 3$, $\vartheta 5$, $\vartheta 8$, $\vartheta 9$, $\vartheta 11$, $\vartheta 16$, $\vartheta 18$, $\vartheta 21$, $\vartheta 22$ and $\vartheta 24$ are indiscernible with respect to the attribute Urinating Often patients $\vartheta 3$, $\vartheta 6$, $\vartheta 9$, $\vartheta 10$, $\vartheta 11$, $\vartheta 12$, $\vartheta 14$, $\vartheta 19$, $\vartheta 20$ and $\vartheta 21$ are indiscernible with respect to attributes Slow Healing and patient $\vartheta 2$, $\vartheta 5$, $\vartheta 8$, $\vartheta 16$, $\vartheta 18$, $\vartheta 22$ and $\vartheta 24$ are indiscernible with respect to attribute Urinating Often,Slow Healing, Weight Loss and Extreme Fatigue .

Hence, for example the attribute Urinating Often generates two elementary sets $\{\vartheta 2$, $\vartheta 3$, $\vartheta 5$, $\vartheta 8$, $\vartheta 9$, $\vartheta 11$, $\vartheta 16$, $\vartheta 18$, $\vartheta 21$, $\vartheta 22$,$\vartheta 24\}$ and $\{\vartheta 1$, $\vartheta 4$, $\vartheta 6$, $\vartheta 7$, $\vartheta 10$, $\vartheta 12$, $\vartheta 13$, $\vartheta 14$, $\vartheta 15$, $\vartheta 17$, $\vartheta 19$, $\vartheta 20$, $\vartheta 23$,$\vartheta 25\}$ where as the attributes Urinating Often and Slow Healing form the elementary set $\{\vartheta 1$, $\vartheta 4$, $\vartheta 6$, $\vartheta 7$, $\vartheta 10$, $\vartheta 12$, $\vartheta 13$, $\vartheta 14$, $\vartheta 15$, $\vartheta 17$, $\vartheta 19$, $\vartheta 20$, $\vartheta 23$, $\vartheta 25\}$, $\{\vartheta 2$, $\vartheta 5$, $\vartheta 8$, $\vartheta 16$, $\vartheta 18$, $\vartheta 22$, $\vartheta 24\}$ and $\{\vartheta 3$, $\vartheta 9$, $\vartheta 11$, $\vartheta 21\}$.Similarly we can define elementary sets generated by any subsets of attributes.

Patient $\vartheta 2$, $\vartheta 18$ and $\vartheta 24$ have Diabetes whereas patient $\vartheta 5$, $\vartheta 8$, $\vartheta 16$ and $\vartheta 22$ do not, and they are indiscernible with respect to the attributes Urinating Often, Slow Healing, Weight Loss, Extreme Fatigue hence diabetes cannot be characterized by the terms of attributes Urinating Often, Slow Healing, Weight Loss, Extreme Fatigue.

Hence, $\vartheta 2$, $\vartheta 5$, $\vartheta 8$, $\vartheta 16$, $\vartheta 18$, $\vartheta 22$ and $\vartheta 24$ are the boundary line cases, which cannot be properly classified in view of the available data.

The remaining patients $\vartheta 1$, $\vartheta 3$, $\vartheta 6$, $\vartheta 9$, $\vartheta 10$, $\vartheta 11$, $\vartheta 12$, $\vartheta 13$, $\vartheta 14$, $\vartheta 17$, $\vartheta 19$, $\vartheta 20$, $\vartheta 21$, $\vartheta 23$ and $\vartheta 25$ display symptoms which enable us to categorize them with certainty as having diabetes in view of the displayed symptoms.

The set of the patients displaying the same symptoms are $\{\vartheta 1$, $\vartheta 2$, $\vartheta 3$, $\vartheta 5$, $\vartheta 6$, $\vartheta 7$, $\vartheta 8$, $\vartheta 9$, $\vartheta 10$, $\vartheta 11$, $\vartheta 12$, $\vartheta 13$, $\vartheta 14$, $\vartheta 15$, $\vartheta 16$, $\vartheta 17$, $\vartheta 18$, $\vartheta 19$, $\vartheta 20$, $\vartheta 21$, $\vartheta 22$, $\vartheta 23$, $\vartheta 24$, $\vartheta 25\}$.

Out of which the set of patients having diabetes are $\{\vartheta 1$, $\vartheta 3$, $\vartheta 6$, $\vartheta 9$, $\vartheta 10$, $\vartheta 11$, $\vartheta 12$, $\vartheta 13$, $\vartheta 14$, $\vartheta 17$, $\vartheta 19$, $\vartheta 20$, $\vartheta 21$,$\vartheta 23$,$\vartheta 25\}$.

Thus, the boundary-line cases are patients $\{\vartheta 2$, $\vartheta 5$, $\vartheta 8$, $\vartheta 16$, $\vartheta 18$, $\vartheta 22$, $\vartheta 24\}$.

Similarly $\{\vartheta 4$, $\vartheta 7$, $\vartheta 15\}$ does not have diabetes and $\{\vartheta 2$, $\vartheta 5$, $\vartheta 8$, $\vartheta 16$, $\vartheta 18$, $\vartheta 22$, $\vartheta 24\}$ cannot be excluded as having diabetes.Thus, the lower approximation of this set is $\{\vartheta 4$, $\vartheta 7$, $\vartheta 15\}$. whereas the upper approximation of this set is $\{\vartheta 2$, $\vartheta 4$, $\vartheta 5$, $\vartheta 7$, $\vartheta 8$, $\vartheta 15$, $\vartheta 16$, $\vartheta 18$, $\vartheta 22$, $\vartheta 24\}$.

Now the boundary conditions are found by finding the difference between the upper and the lower approximations ,(i.e) {ϑ2, ϑ4, ϑ5, ϑ7, ϑ8, ϑ15, ϑ16, ϑ18, ϑ22, ϑ24}-{ϑ4, ϑ7, ϑ15}.

Therefore the boundary region of the set is {ϑ2, ϑ5, ϑ8, ϑ16, ϑ18, ϑ22, ϑ24}.

4. Conclusions

In this paper, we have discussed one of the applications based on Rough set theory.With the lower and upper approximations of the decision concepts, we can find a boundary-line for diabetes from the information table of the patients with the same symptoms.

References

[1] ADAMSON: *A General Topology Workbook*, Boston, MA: Birkhäuser, 1996.

[2] Z.PAWLAK: *Rough sets*, International Journal of Computer and Information Science, **11** (1982), 341-356.

[3] Z.PAWLAK: *Rough Sets -Theoretical Aspects of Reasoning about Data*, Kluwer Academyc Publisher, Boston, London, Dordrecht, 1991.

[4] A. SKOWRON, C. RAUSZER: *The Discernibility matrices and functions in information systems,Decision Support Hand book of Applications and Advances of the Rough Set Theory*, R.Slowinski (ed.), Intelligent Kluwer Academic Publishers, Dordrecht, (1992), 311-362.

PG AND RESEARCH DEPARTMENT OF MATHEMATICS
KONGUNADU ARTS AND SCIENCE COLLEGE, G.N.MILLS (PO)
COIMBATORE-641 029, TAMIL NADU, INDIA
E-mail address: sarjunraj@kongunaducollege.ac.in

PG AND RESEARCH DEPARTMENT OF MATHEMATICS
KONGUNADU ARTS AND SCIENCE COLLEGE, G.N.MILLS (PO)
COIMBATORE-641 029, TAMIL NADU, INDIA
E-mail address: vigneshmaths@kongunaducollege.ac.in