Int. J. Math. And Appl., 6(1–E)(2018), 1165–1174
ISSN: 2347-1557
Available Online: http://ijmaa.in/

International Journal of Mathematics And its Applications

Analyzing of Ebola Disease Using Rough Set Theory

Y. Preethi Ceon[1,*]

1 Department of Mathematics, KG College of Arts and Science, Coimbatore, Tamilnadu, India.

Abstract: The concept of rough set theory in data mining is discussed. Using data mining we diagonize Ebola disease. The pre n post processing stage has been discussed. The lower & upper approximation of set are interior and closure operation is generated for Ebola. Analysis of the artibutes is given in a table.

Keywords: Data mining, closure, interior.

1. Introduction-Data Mining

Data Mining generally refers to collection of data. It is a Knowledge discovering Database which consist of three stages:

- Preprocessing Stage
- Data Mining
- Post Processing

Preprocessing Stage: The preprocessing stage understands the functions related to the reception, the organization and to the treatment of data. It's objective the preparation of the data for the following stage of the data mining.

Data Mining: The data mining stage defines the techniques and the algorithms to be used in Processed data. Three ways of Processing data are given by: First, Example based on Rough set (ERS) uses only the symbolic attributes. For, the inconsistent input data, (ERS)computes lower a Upper approximations of the given concept is called Possible.

Second, Examples Module version 2 (EM2) uses both symbolic and numerical attributes. The original algorithm (EM2) needs discretization, a pre processing to deal with numerical attributes. Discretization is the process of converting numerical attributes with intervals as values. EM2 treats all attributes as symbolic, thus producing too specific rules when the input data are not discretized.

Third Modified Examples Module Version 2 (MEM2), a new algorithm extends EM2 capabilities by including rules from data with symbolic and numerical attributes and including data with missing attributes values.

Post Processing Stage: In the post processing stage, the treatment of knowledge is obtained during the data mining stage. This stage is not always necessary; however, it allows the possibility of validation of the usefulness of the discovered knowledge. As an application, we use Rough set tools and data mining technique to diagonize the Ebola disease. The

* *E-mail: preethiceon.y@kgcas.com*

symptoms of ebola will be joint pain, brain damage and bleeding which will cause Ebola. Now, We will deal with several patients data and we will analysis the data using a Rough Set approach for the elimination of reduct and data and the development of a set of rules that it can aid the doctor in the elaboration of the diagnosis

2. Ebola Information Table

In this, we are given with several patients data set with possible Ebola symptoms. Here we consider the information system of the form $S = (U, A)$, where U is the collection of 15 patients and A is the union of conditional and decision attributes.

Patient	Conditional Attributes			Decision Attribute
	Joint pain	Brain damage	Bleeding	Ebola
P1	Yes	Yes	Internal bleeding	Yes
P2	Yes	No	External bleeding	No
P3	Yes	Yes	No bleeding	No
P4	No	Yes	External bleeding	Yes
P5	Yes	No	Internal bleeding	No
P6	Yes	No	External bleeding	No
P7	Yes	No	Internal bleeding	Yes
P8	No	No	External bleeding	No
P9	Yes	Yes	Internal bleeding	Yes
P10	Yes	Yes	External bleeding	Yes
P11	No	Yes	Internal bleeding	Yes
P12	No	Yes	External bleeding	Yes
P13	No	No	Internal bleeding	Yes
P14	No	No	External bleeding	No
P15	No	No	No bleeding	No

Table 1. **Patients with respective symptoms**

Here the set B are the objects or registrations of the system,

$$B = \{P1,\ P2,\ P3,\ P4,\ P5,\ P6,\ P7,\ P8,\ P9,\ P10,\ P11,\ P12,\ P13,\ P14,\ P15\},$$

Where the conditional attributes is represented by $C = \{Joint\ pain,\ Brain\ damage,\ Bleeding\}$ and the decision attribute is represented by $D = \{Ebola\}$. The Table 1, can be shown in relation to the function of nominal values of considered attributes, in the Table 2:

	Attributes	Nominal values
Conditional Attributes	Joint pain	Yes, No
	Brain damage	Yes, No
	Bleeding	No bleeding, External bleeding, Internal bleeding
Decision Attributes	Ebola	Yes, No

Table 2. **Shows Nominal Values of Attributes**

3. Indiscernibility Relation

Indiscernibility Relation is the relation between two objects or more, where all the values are identical in relation to a subset of considered attributes. In Table 1.1, it can be observed that the set is composed of attributes that are directly related to patient's symptoms, where $C = \{Joint\ pain,\ \ Brain\ damge,\ \ Bleeding\}$, the indiscernibility relation is represented as

$IND(C)$. When Table 1 is broken down it can be easily seen that indiscernibility relation is in relationship to conditional attributes:

(1). The Joint pain attribute generates two indiscernibility elementary sets:

$$IND\,([Joint\ pain]) = \{\{P4,\ P8,\ P11,\ P12,\ P13,\ P14,\ P15\}, \{P1,\ P2,\ P3,\ P5,\ P6,\ P7,\ P9,\ P10\}\}$$

Patient	Conditional Attributes			Decision Attribute
	Joint pain	Brain damage	Bleeding	Ebola
P4	No	Yes	External bleeding	Yes
P8	No	No	External bleeding	No
P11	No	Yes	Internal bleeding	Yes
P12	No	Yes	External bleeding	Yes
P13	No	No	Internal bleeding	Yes
P14	No	No	External bleeding	No
P15	No	No	No bleeding	No
P1	Yes	Yes	Internal bleeding	Yes
P2	Yes	No	External bleeding	No
P3	Yes	Yes	No bleeding	No
P5	Yes	No	Internal bleeding	No
P6	Yes	No	External bleeding	No
P7	Yes	No	Internal bleeding	Yes
P9	Yes	Yes	Internal bleeding	Yes
P10	Yes	Yes	External bleeding	Yes

Table 3. **Organize in relation with Joint pain**

(2). The Brain damage attribute generates two indiscernibility elementary sets:

$$IND(\{Brain\ damage\}) = \{\{P2,\ P5,\ P6,\ P7,\ P8,\ P13,\ P14,\ P15\}, \{P1,\ P3,\ P4,\ P9,\ P10,\ P11,\ P12\}\}$$

Patient	Conditional Attributes			Decision Attribute
	Joint pain	Brain damage	Bleeding	Ebola
P2	Yes	No	External bleeding	No
P5	Yes	No	Internal bleeding	No
P6	Yes	No	External bleeding	No
P7	Yes	No	Internal bleeding	Yes
P8	No	No	External bleeding	No
P13	No	No	Internal bleeding	Yes
P14	No	No	External bleeding	No
P15	No	No	No bleeding	No
P1	Yes	Yes	Internal bleeding	Yes
P3	Yes	Yes	No bleeding	No
P4	No	Yes	External bleeding	Yes
P9	Yes	Yes	Internal bleeding	Yes
P10	Yes	Yes	External bleeding	Yes
P11	No	Yes	Internal bleeding	Yes
P12	No	Yes	External bleeding	Yes

Table 4. **Organize in relation with brain damage**

(3). The attribute bleeding generates three indiscernibility elementary sets:

$$IND\,([bleeding]) = \{\{P3,\ P15\},\ \{P2,\ P4,\ P6,\ P8,\ P10,\ P12,\ P14\}, \{P1,\ P5,\ P7,\ P9,\ P11,\ P13\}$$

Patient	Conditional Attributes			Decision Attribute
	Joint pain	Brain damage	Bleeding	Ebola
P3	Yes	Yes	No bleeding	No
P15	No	No	No bleeding	No
P2	Yes	No	External bleeding	No
P4	No	Yes	External bleeding	Yes
P6	Yes	No	External bleeding	No
P8	No	No	External bleeding	No
P10	Yes	Yes	External bleeding	Yes
P12	No	Yes	External bleeding	Yes
P14	No	No	External bleeding	No
P1	Yes	Yes	Internal bleeding	Yes
P5	Yes	No	Internal bleeding	No
P7	Yes	No	Internal bleeding	Yes
P9	Yes	Yes	Internal bleeding	Yes
P11	No	Yes	Internal bleeding	Yes
P13	No	No	Internal bleeding	Yes

Table 5. Organize in relation with Bleeding

4. Approximations

The lower and upper approximations of a set are interior and closure operations in a topology generated by a indiscernibility relation. The approximations concepts are applied in the Table 1.1 are shown below:

Patient	Conditional Attributes			Decision Attribute
	Joint pain	Brain damage	Bleeding	Ebola
P2	Yes	No	External bleeding	No
P3	Yes	Yes	No bleeding	No
P5	Yes	No	Internal bleeding	No
P6	Yes	No	External bleeding	No
P8	No	No	External bleeding	No
P14	No	No	External bleeding	No
P15	No	No	No bleeding	No
P1	Yes	Yes	Internal bleeding	Yes
P4	No	Yes	External bleeding	Yes
P7	Yes	No	Internal bleeding	Yes
P9	Yes	Yes	Internal bleeding	Yes
P10	Yes	Yes	External bleeding	Yes
P11	No	Yes	Internal bleeding	Yes
P12	No	Yes	External bleeding	Yes
P13	No	No	Internal bleeding	Yes

Table 6. Organize in relation with Ebola

(a). **Lower Approximation Set $B_*(X)$**

(1). Lower Approximation set $B_*(X)$ of the patients that are definitely have Ebola are identified as $B_*(X) = \{P1,\ P4,\ P9,\ P10,\ P11,\ P12,\ P13\}$.

(2). Lower Approximation set $B_*(X)$ of patients that certain have not Ebola are identified as $B_*(X) = \{P2,\ P3,\ P6,\ P8,\ P14,\ P15\}$.

(b). **Upper Approximation Set $B^*(X)$**

(1). Upper Approximation set $B^*(X)$ of the patients that possibly have Ebola are identified as $B^*(X) = \{P1,\ P4,\ P7,\ P9,\ P10,\ P11,\ P12,\ P13\}$.

(2). Upper Approximation set $B^{*}(X)$ of the patients that possibly have not Ebola are identified as $B^{*}(X) = \{P2,\ P5,\ P6,\ P8,\ P14,\ P15\}$.

(c). **Boundary Region (BR)**

(1). Boundary Region of the patients that not have Ebola are identified as: $BR = \{P2,\ P3,\ P5,\ P6,\ P8,\ P14,\ P15\} - \{P2,\ P3,\ P6,\ P8,\ P14,\ P15\} = \{P5\}$.

(2). Boundary Region the set of the patients that have Ebola are identified as: $BR = \{P1,\ P4,\ P9,\ P10,\ P11,\ P12,\ P13\} - \{P1,\ P4,\ P7,\ P9,\ P10,\ P11,\ P12,\ P13\} = \{P7\}$.

Result: Boundary region, the set constituted by elements P7 and P5, which cannot be classified because they possess the same characteristics, but conclusions differ in the decision attribute.

5. Quality of Approximation

The two coefficients of quality of approximation are:

Imprecision coefficient:

(1). for the patients with possibility of having ebola $aB(X) = \frac{7}{8}$.

(2). for the patients with possibility of not having ebola $aB(X) = \frac{8}{12}$.

Quality Coefficient of upper and lower approximation:

(1). $aB(B^{*}(X)) = \frac{8}{15}$, for the patients having the possibility ofebola;

(2). $aB(B^{*}(X)) = \frac{6}{15}$, for the patients not having the Ebola;

(3). $aB(B_{*}(X)) = \frac{7}{15}$, for the patients that have ebola;

(4). $aB(B_{*}(X)) = \frac{6}{15}$, for the patients not having Ebola.

Result:

1. Patients with Ebola: 47%
2. Patients that don't have Ebola: 40%
3. 13% of patients P5 and P11 cannot be classified neither with Ebola nor without Ebola. Since characteristics of all attributes are the same, with only the decision attribute- Ebola not being identical and generates an inconclusive diagnosis for Ebola.

6. Data Reduction in Information System

A reduct is a set of necessary minimum data, since the original proprieties of the system or information table are maintained. Therefore, the reduct must have the capacity to classify objects, without altering the form of representing the knowledge. Let us once again consider Table 1:

(1). **Verification of inconclusive data**

Step 1: Analysis of data contained in Table 1 shows that information is inconclusive, because the values of conditional attributes same and the value of decision attribute is different.

Conclusion of Step 1: The symptoms of patient P5 and patient P7 will be excluded, since they possess equal values of conditions attributes and different decision value. Therefore, the data of patient P5 and patient P7 will be excluded from Table 1.

(2). **Verification of equivalent information**

Step 2: Analysis of data contained in Table 1 shows that it has equivalent information.

P8	No	No	External bleeding	No
P14	No	No	External bleeding	No

P4	No	Yes	External bleeding	Yes
P12	No	Yes	External bleeding	Yes

P1	Yes	Yes	Internal bleeding	Yes
P9	Yes	Yes	Internal bleeding	Yes

P2	Yes	No	External bleeding	No
P6	Yes	No	External bleeding	No

Conclusion of Step 2: The Table 1 has it reduced data presented in a revised version in the Table 7 shown below:

Patient	Conditional Attributes			Decision Attribute
	Joint pain	Brain damage	Bleeding	Ebola
P3	Yes	Yes	No bleeding	No
P6	Yes	No	External bleeding	No
P8	No	No	External bleeding	No
P9	Yes	Yes	Internal bleeding	Yes
P10	Yes	Yes	External bleeding	Yes
P11	No	Yes	Internal bleeding	Yes
P12	No	Yes	External bleeding	Yes
P13	No	No	Internal bleeding	Yes
P14	No	No	External bleeding	No
P15	No	No	No bleeding	No

Table 7. **Revised version of Table 1**

Step 3: Analysis of each condition attributes with the attributes set.

Patient	Conditional Attribute	Decision Attribute
	Joint pain	Ebola
P3	Yes	No
P6	Yes	No
P8	No	No
P9	Yes	Yes
P10	Yes	Yes
P11	No	Yes
P12	No	Yes
P13	No	Yes
P14	No	No
P15	No	No

Table 8. **Shows the analysis of attribute Joint pain**

Patient	Conditional Attribute	Decision Attribute
	Brain damage	Ebola
P3	Yes	No
P6	No	No
P8	No	No
P9	Yes	Yes
P10	Yes	Yes
P11	Yes	Yes
P12	Yes	Yes
P13	No	Yes
P14	No	No
P15	No	No

Table 9. **Shows the analysis of brain damage**

Patient	Conditional Attribute	Decision Attribute
	Bleeding	Ebola
P3	No bleeding	No
P6	External bleeding	No
P8	External bleeding	No
P9	Internal bleeding	Yes
P10	External bleeding	Yes
P11	Internal bleeding	Yes
P12	External bleeding	Yes
P13	Internal bleeding	Yes
P14	External bleeding	No
P15	No bleeding	No

Table 10. **Shows the analysis of Bleeding**

Conclusion of the above analysis is that no data is excluded. Given analysis of condition attributes in Table 7, it can be observed that the same data exists in proceeding tables.

- Analysis of attributes Joint pain and brain damage in Table 7 is shown in Table 11.

Patient	Conditional Attributes		Decision Attribute
	Joint pain	Brain damage	Ebola
P3	Yes	Yes	No
P6	Yes	No	No
P8	No	No	No
P9	Yes	Yes	Yes
P10	Yes	Yes	Yes
P11	No	Yes	Yes
P12	No	Yes	Yes
P13	No	No	Yes
P14	No	No	No
P15	No	No	No

Table 11.

Result analysis:

Patient	Conditional Attributes		Decision Attribute
	Joint pain	Brain damage	Ebola
P3	Yes	Yes	No
P6	Yes	No	No
P8	No	No	No
P9	Yes	Yes	Yes
P10	Yes	Yes	Yes
P11	No	Yes	Yes
P13	No	No	Yes

Table 12.

Analysis of Attributes Joint pain and Bleeding is Shown in Table 13

Patient	Conditional Attributes		Decision Attributes
	Joint pain	Bleeding	Ebola
P3	Yes	No bleeding	No
P6	Yes	External bleeding	No
P8	No	External bleeding	No
P9	Yes	Internal bleeding	Yes
P10	Yes	External bleeding	Yes
P11	No	Internal bleeding	Yes
P12	No	External bleeding	Yes
P13	No	Internal bleeding	Yes
P14	No	External bleeding	No
P15	No	No bleeding	No

Table 13.

Result analysis:

Patient	Conditional Attributes		Decision Attributes
	Joint pain	Bleeding	Ebola
P3	Yes	No bleeding	No
P6	Yes	External bleeding	No
P8	No	External bleeding	No
P9	Yes	Internal bleeding	Yes
P10	Yes	External bleeding	Yes
P11	No	Internal bleeding	Yes
P12	No	External bleeding	Yes
P15	No	No bleeding	No

Table 14.

Analysis of attributes brain damage and Bleeding is shown in Table 15.

Patient	Conditional Attributes		Decision Attribute
	Brain damage	Bleeding	Ebola
P3	Yes	No bleeding	No
P6	No	External bleeding	No
P8	No	External bleeding	No
P9	Yes	Internal bleeding	Yes
P10	Yes	External bleeding	Yes
P11	Yes	Internal bleeding	Yes
P12	Yes	External bleeding	Yes
P13	No	Internal bleeding	Yes
P14	No	External bleeding	No
P15	No	No bleeding	No

Table 15.

Result analysis:

Patient	Conditional Attributes		Decision Attribute
	Brain damage	Bleeding	Ebola
P3	Yes	No bleeding	No
P6	No	External bleeding	No
P9	Yes	Internal bleeding	Yes
P10	Yes	External bleeding	Yes
P11	Yes	Internal bleeding	Yes
P13	No	Internal bleeding	Yes
P15	No	No bleeding	No

Table 16.

Step 4: Verification of equivalent (intersection) data in the Tables 12, 14 and 16 corresponds to the data which is the element of reduction formation in relation to Table 7.

Patient	Conditional Attributes			Decision Attribute
	Joint pain	Brain damage	Bleeding	Ebola
P3	Yes	Yes	No bleeding	No
P6	Yes	No	External bleeding	No
P9	Yes	Yes	Internal bleeding	Yes

Table 17.

7. Decision Rules

With the information reduct shown above, it can be generated the necessary decision rules for aid to the Ebola disease. The rules are presented to proceed:

Rule-1

R1: If patient

Joint pain = Yes

Brain damage =Yes

Bleeding =No bleeding

Then Ebola =No

Rule-2

R2: If patient

Joint pain = Yes

Brain damage =No

Bleeding =External bleeding

Then Ebola =No

Rule-3

R3: If patient

Joint pain = Yes

Brain damage =Yes

Bleeding = Internal bleeding

Then Ebola =Yes

8. Conclusion

From the analysis of the Ebola we come to know that if all three symptoms are found, patient have Ebola. Otherwise, The patient have some other disease other than Ebola.

References

[1] Adam Grabowski, *Basic Properties of Rough Sets and Rough Membership Function*, Formalized Mathematics, 12(1)(2004), 21-28.

[2] J. Bazan, A. Skrowron and P. Synak, *Discovery of decision rules from experimental data*, in: T.Y. Lin and A.M. Wildberger, The Third International Workshop on Rough Sets and Soft Computing Proceedings (RSSC'94), San Jose State University, San Jose, California, USA, November 10-12, 276-279.

[3] T. Beaubouef and F.E. Petry, *A rough set model for relational databases*, in: W. Ziarko (ed.), Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedingsof the International Workshop on Rough Sets and Knowledge Discovery (RSKD'93), Banff, Alberta, Canada, October 12-15(1993), Springer-Verlag, Berlin, 100-107.

[4] A. Czyźewski and A. Kaczmarek, *Speaker-independent recognition of isolated words using rough sets*, in: P.P. Wang (ed.), Second Annual Joint Conferenceon Information Sciences PROCEEDINGS, September 28-October 1(1995), Wrightsville Beach, North Carolina, USA, 397-400.

[5] J.W. Grzymalℓa-Busse, *Knowledge acquisition under uncertainty-A rough set approach*, Journal of Intelligent & Robotic Systems, 1(1)(1988), 3-16.

[6] Gaolin Li and X. Luoshan, *Lattice Structures of Probabilistic Rough Sets*, Communications in Information Science and Management Engineering, 3(6)(2013), 270-279.

[7] Kanchan S. Tiwari, Ashwin G. Kothari and Avinash G. Keskar, *Reduct Generation from Binary Discernibility Matrix: An Hardware Approach*, International Journal of Future Computer and Communication, 1(3)(2012).

[8] Y. Preethi Ceon, *Analysis of chikungunya by process of data reduction*, International Journal of Science & Engineering Devolpment Research, 2(5)(2017).