ANALYSIS OF CHIKUNGUNYA BY PROCESS OF DATA REDUCTION

Y. Preethi Ceon

Assistant professor
Department of Mathematics
KG College of Arts and Science
Coimbatore-641035,Tamil Nadu, India.

ABSTRACT: In this paper, the analysis of chikungunya using the method of data mining has been done. Using the method we are able to deduct the percentage of the effect of chikungunya in human body and we can hence go forth for further treatment in order to stabilize the condition of patient.

1. INTRODUCTION-DATA MINING

Data Mining generally refers to collection of data. It is a Knowledge discovery in Database which consists of three stages:

- Pre processing Stage
- Data Mining
- Post Processing

PRE PROCESSING STAGE

The pre processing stage will help to understand the functions related to the reception, the organization and to the treatment of data. It's objective is the preparation of the data for the data mining stage.

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:

I. Example based on Rough set (ERS) uses only the symbolic attributes. For, the inconsistent input data, (ERS) computes lower and upper approximation of all concepts. Rules induced from the lower approximations of the given concept is called Certain. Rules induced from the Upper approximations of the given concept is called Possible.

II. Second, Examples Module version2 (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.

III. Modified Examples Module Version2 (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, Using Rough set tools and data mining technique to diagonize the chikungunya disease. Chikungunya is the type of skin disease. The symptoms of chikungunya will be scalp in the hair, blood red itching skin and muscular pain which will cause chikungunya. Now, Dealing with several patients data and will analyse the data using a Rough Set approach for the elimination of reductant data and the development of a set of rules that can aid the doctor in the elaboration of the diagnosis. Here, given the discrete data.

1.2 CHIKUNGUNYA INFORMATION TABLE

In this section, we are given with several patients data set with possible chikungunya symptoms. Here we consider the information system of the form $S = (U, A)$ where$U$ is the collection of 10 patients and A is the union of conditional and decision attributes.

PATIENT	CONDITIONAL ATTRIBUTES			DECISION ATTRIBUTES
	Body pain	**Fever**	**Muscular pain**	**Chikungunya**
P1	Yes	Yes	Severe Pain	Yes
P2	Yes	No	Moderate Pain	No
P3	Yes	Yes	No Pain	No
P4	No	Yes	Moderate Pain	Yes
P5	Yes	No	Severe Pain	No
P6	Yes	No	Moderate Pain	No
P7	Yes	No	Severe Pain	Yes
P8	No	No	Moderate Pain	No
P9	Yes	Yes	Severe Pain	Yes
P10	Yes	Yes	Moderate Pain	Yes

Table 1.1

The above table shows patients with respective symptoms.Here the set B are the objects or registrations of the system,

$$B = \{P1, P2, P3, P4, P5, P6, P7, P8, P9, P10,\}$$

where the conditional attributes is represented by $C = \{bodypain,\ fever,\ muscularpain\}$ and the decision attribute is represented by$D = \{Chikungunya\}$. The Table 3.1, can be shown in relation to the function of nominal values of considered attributes, in the Table 1.2:

	Attributes	Nominal values
Conditional Attributes	Body pain	Yes, No
	Fever	Yes, No
	Muscular pain	No Pain, Moderate Pain, Severe Pain
Decision Attributes	Chikungunya	Yes, No

Table 1.2

The above table shows Nominal Values of Attributes.

1.3 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 3.1, it can be observed that the set is composed of attributes that are directly related to patient's symptoms, where $C = \{bodypain, fever, andmuscularpain\}$,the Indiscernibility relation is represented as $IND(C)$. When Table 1.1 is broken down it can be easily seen that Indiscernibility relation is in relationship to conditional attributes:

1. The Body pain attribute generates two Indiscernibility elementary sets: $IND([Bodypain]) = \{\{P4, P8,\},\ \{P1, P2, P3, P5, P6, P7, P9, P10\}\}$

PATIENT	CONDITIONAL ATTRIBUTES			DECISION ATTRIBUTE
	Body pain	Fever	Muscular Pain	Chikungunya
P4	No	Yes	Moderate Pain	Yes
P8	No	No	Moderate Pain	No
P1	Yes	Yes	Severe Pain	Yes
P2	Yes	No	Moderate Pain	No
P3	Yes	Yes	No Pain	No
P5	Yes	No	Severe Pain	No
P6	Yes	No	Moderate Pain	No
P7	Yes	No	Severe Pain	Yes
P9	Yes	Yes	Severe Pain	Yes
P10	Yes	Yes	Moderate Pain	Yes

Table 1.3

The above table organize in relation with body pain

2. The fever attribute generates two indiscernibility elementarysets:

$$IND(\{fever\}) = \{\{P2, P5, P6, P7, P8, \},$$

$$\{P1, P3, P4, P9, P10, \}$$

PATIENT	CONDITIONAL ATTRIBUTES			DECISION ATTRIBUTE
	Body pain	Fever	Muscular Pain	Chikungunya
P2	Yes	No	Moderate Pain	No
P5	Yes	No	Severe Pain	No
P6	Yes	No	Moderate Pain	No
P7	Yes	No	Severe Pain	Yes
P8	No	No	Moderate Pain	No
P1	Yes	Yes	Severe Pain	Yes
P3	Yes	Yes	No Pain	No
P4	No	Yes	Moderate Pain	Yes
P9	Yes	Yes	Severe Pain	Yes
P10	Yes	Yes	Moderate Pain	Yes

Table 1.4

The above table organize in relation with fever.

3. The attribute muscular pain generates three in discernibility elementarysets:

$$IND([muscularpain]) = \{\{P3, \}, \{P2, P4, P6, P8, P10, \},$$

$$\{P1, P5, P7, P9, \}$$

PATIENT	CONDITIONAL ATTRIBUTES			DECISION ATTRIBUTE
	Body pain	Fever	Muscular Pain	Chikungunya
P3	Yes	Yes	No Pain	No
P2	Yes	No	Moderate Pain	No
P4	No	Yes	Moderate Pain	Yes
P6	Yes	No	Moderate Pain	No
P8	No	No	Moderate Pain	No
P10	Yes	Yes	Moderate Pain	Yes
P1	Yes	Yes	Severe Pain	Yes
P5	Yes	No	Severe Pain	No
P7	Yes	No	Severe Pain	Yes
P9	Yes	Yes	Severe Pain	Yes

Table 1.5

The above table organize in relation with Muscular Pain.

1.4 APPROXIMATIONS

The lower and the 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
	Body pain	Fever	Muscular Pain	Chikungunya
P2	Yes	No	Moderate Pain	No
P3	Yes	Yes	No Pain	No
P5	Yes	No	Severe Pain	No
P6	Yes	No	Moderate Pain	No
P8	No	No	Moderate Pain	No
P1	Yes	Yes	Severe Pain	Yes
P4	No	Yes	Moderate Pain	Yes
P7	Yes	No	Severe Pain	Yes
P9	Yes	Yes	Severe Pain	Yes
P10	Yes	Yes	Moderate Pain	Yes

Table 1.6

The above table organize in relation with chikungunya.

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

- Lower Approximation set$B_*(X)$ of the patients that are definitely have chikungunya are identified as $B_*(X) = \{P1, P4, P9, P10,\}$.
- Lower Approximation set $B_*(X)$ of patients that certain have not chikungunya are identified as$B_*(X) = \{P2, P3, P6, P8,\}$.

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

- Upper Approximation set $B^*(X)$ of the patients that possiblyhave chikungunya are identified as $B^*(X) = \{P1, P4, P7, P9, P10,\}$.
- Upper Approximation set$B^*(X)$ of the patients that possiblyhave not chikungunya are identified as $B^*(X) = \{P2, P5, P6, P8,\}$.

c. Boundary Region (BR)

- Boundary Region of the patients that not have chikungunya are identified as: $BR = \{P2, P3, P5, P6, P8,\} - \{P2, P3, P6, P8,\} = \{P5\}$.
- Boundary Region the set of the patients that have chikungunya are identified as:

$$BR = \{P1, P4, P9, P10,\} - \{P1, P4, P7, P9, P10\} = \{P7\}$$

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.

1.5 QUALITY OF APPROXIMATION

The two coefficients of quality of approximation are:

Imprecision coefficient:

- for the patients with possibility of having chikungunya$\alpha B(X) = 5/7$.
- for the patients with possibility of not having chikungunya$\alpha B(X) = 4/8$.

Quality Coefficient of upper and lower approximation:

- $\alpha B(B^*(X)) = 5/10$, for the patients having the possibility of chikungunya;
- $\alpha B(B^*(X)) = 4/10$, for the patients not having the chikungunya;
- $\alpha B(B_*(X)) = 4/10$, for the patients that have chikungunya;
- $\alpha B(B_*(X)) = 4/10$, for the patients not having chikungunya.

CONCLUSION

We have analysed the chikungunya diseases using data reduction in the following process

1. Patients with chikungunya: 47%
2. Patients that don't have chikungunya: 40%
3. 13% of patients P5 and P10 cannot be classified neither with chikungunya nor without chikungunya, since the characteristics of all attributes are the same, with only the decision attribute- chikungunya not being identical and generates an in conclusive diagnosis for chikungunya.

REFERENCES:

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

[2] Bazan, J., Skrowron, A., and Synak, P., "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] Beaubouef, T., and Petry, F.E., (1993), "A rough set model for relational databases", in: W. Ziarko (ed.), Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery (RSKD'93), Banff, Alberta, Canada, October 12–15, Springer-Verlag, Berlin, 100–107.

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

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