International Journal of Bio-Science and Bio-Technology
Vol.6, No.5 (2014), pp.143-152
http://dx.doi.org/10.14257/ijbsbt.2014.6.5.14

Application of IF Rough Set on Knowledge Towards Malaria of Rural Tribal Communities in Tripura

Chhaya Gangwal[1], R. N. Bhaumik[2] and Shishir Kumar[3]

[1] Research Scholar, Tripura University
c_hhaya@hotmail.com
[2]Emeritus Fellow (UGC) and Professor of Mathematics(Rtd.)
bhaumik_r_n@yahoo.co. in
[3]Assistant Professor (Biostatistics), Department of Community Medicine,
Agartala Government Medical College, Agartala, Tripura
shishirpunia@gmail.com

Abstract

Handling uncertainty and impreciseness of knowledge appears to be a challenging task in Information Systems. Intuitionistic fuzzy (IF) set theory and rough set theory enhances databases by allowing it for the management of uncertainty and impreciseness. This paper presents a new efficient query optimization technique for the multi-valued or imprecise IF rough database. The usefulness of this technique was illustrated on malaria knowledge from the rural tribal communities of Tripura where most of the information is multi-valued and imprecise. Then the querying about knowledge on malaria is executed into SQL server to make the implementation of IF rough data querying simpler.

Keywords: *Intuitionistic Fuzzy Set, Rough Set, Relational Database, IF Rough Relational Database*

1. Introduction

Many real world applications and systems deal with imprecise, uncertain or multi-valued data. For such systems, we need information management systems that provide support for managing this imprecise and uncertain data. Significant work has been done in incorporating uncertainty management in relational databases (RDBs) [3] using theories such as fuzzy set [10], rough set [9] and intuitionistic fuzzy (IF) set [1] to name a few. Each of the theories has some advantages in modeling some type of uncertainty over another.

At present, there are only a few researches on the theory architecture of IF and rough relational database. Currently, the study of rough data querying on simple select-querying are proposed by [7, 8] for multi-valued attributes, whereas, there are many research issues regarding the management of imprecise data in database environment. To manage impreciseness in relational databases IF set theory finding wide usefulness.

Recently [5], we presented an IF rough relational database model and described the IF rough SQL used for querying in IF rough databases and applied [6] this model on diabetic patient databases. In this paper a new efficient query optimization technique for the multi-valued or imprecise IF rough database on malaria knowledge from the rural tribal communities of Tripura is presented. This methodology is perceived by using encoding and normalizing techniques.

ISSN: 2233-7849 IJBSBT
Copyright ⓒ 2014 SERSC

2. Preliminaries

2.1. Intuitionistic Fuzzy (IF) set [1]

An IF set A in a nonempty set X is A = {(x, $\mu_A(x)$, $\nu_A(x)$) : x ∈ X }, where $\mu_A(x)$ and $\nu_A(x)$ are functions from X to I = [0, 1] such that $0 \le \mu_A(x) + \nu_A(x) \le 1$, ∀ x ∈ X. The numbers $\mu_A(x)$ and $\nu_A(x)$ represent the degree of membership and degree of non-membership for each element x ∈ X to A respectively. The quantity $\pi_A(x) = 1 - (\mu_A(x) + \nu_A(x))$ is called the degree of indeterminacy or hesitation of the element x ∈ X to the IF set A.

2.2. Rough Set [9]

Let *R* be an indiscernibility relation on universal set *U*. The pair *S* = (*U*, *R*) is called a Pawlak approximation space. Then for any non-empty subset *X* of *U*, the sets $\underline{R}X = \{x \in U : [x]_R \subseteq X\}$ and $\overline{R}X = \{x \in U : [x]_R \cap X \neq \phi\}$ are respectively, called the lower and the upper approximations of *X* in *S*. The set approximation $\underline{R}X$, $(U - \overline{R}X)$ and $(\overline{R}X - \underline{R}X)$ are described as R-positive region, R-negative region and R-boundary region respectively, where $[x]_R$ denotes the equivalence class of the relation *R* containing the element *x*. *X* is said to be definable set, if $\overline{R(X)} = \underline{R(X)}$. Otherwise *X* is said to be rough set. Rough relational database [2] is a special kind of multi-valued information system S. Let *S* = (U, A, D, R), U is the set of all the tuples, A is the attribute set, D is the domains of attribute sets, and R is the equivalence classes on the D. For any attribute A_i ∈ A, D_{Ai} is the domain of A_i, R_{Ai} is the equivalence class of attribute A_i, a tuple r ∈ U, $r(A_i)$ is the tuple r's value on attribute A_i, and $r(A_i) \subseteq D_{Ai}$.

3. IF Rough Relational Database Model [5, 6]

In this model, a tuple t_i takes the form (d_{i1}, d_{i2}, . . , d_{im}, $d_{i[\mu,\nu]}$) where d_{ij} is a domain value of a particular domain set D_j and $d_{i[\mu,\nu]} \in [0, 1]$, the domain for IF membership and non-membership values denoted as $d_{i[\mu,\nu]} = [d_{i\mu}, d_{i\nu}]$. In the relational database, $d_{ij} \in D_j$. In the IF rough relational database except for the membership and non-membership values $d_{ij} \subseteq D_j$ where $d_{ij} \neq \phi$.

Definition 1 Let $P(D_i)$ be the power set of D_i. An IF rough relation R is a subset of the product set $P(D_1) \times P(D_2) \times \ldots P(D_m) \times D_{[\mu,\nu]}$, where $D_{[\mu,\nu]}$ is the domain for membership and non-membership value of the closed interval [0,1] and $P(D_i) = P(D_i) - \phi$.
Definition 2 Let $t_i = (d_{i1}, d_{i2}, ..., d_{im}, d_{i[\mu,\nu]})$ be an IF rough tuple. An interpretation of t_i is a tuple $\alpha = (a_1, a_2, ..., a_{m,n}, a_{[\mu,\nu]})$ where $a_j \in d_{ij}$ for each domain D_j.

3.1. IF Rough Methodology

Suppose the IF rough relation R and multi-valued or imprecise attribute $a \in A$, D_a is attribute *a*'s domain, R_a is attribute *a*'s equivalence class and $D_{[\mu,\nu]}$ is the domain for membership and non-membership value of the closed interval [0,1].
Then, it can be noted that: $D_a / R_a = \{[x]_{Ra} \mid x \in D_a\} = \{c_1, c_2, \ldots, c_k\}$, and $K = |D_a / R_a|$.
Let v be an arbitrary value of the multi valued attribute *a* and ENCODE be a map function. ENCODE: ENCODE (a, v) → $c_1, c_2, \ldots c_k$, where $v \subseteq D_a$. $c_i = 1$ if $x \in v \wedge x \in c_i$, otherwise $c_i = 0$ and for imprecise attribute *a*, removing all the imprecise attributes from the relation R,

 Copyright ⓒ 2014 SERSC

create one separate table for each imprecise attribute, and put all attributes in the primary key with $D_{[\mu,\nu]}$.
Based on the above methodology, we get the following algorithm.

3.2. IF Rough Algorithm

Let the select-condition be "a = v", a is an attribute and its domain is D_a and *a_code* is the encoding filed of a and a_μ and a_ν are the membership and non-membership values of a and v is an arbitrary value, where $v \subseteq D_a$.

1. Calculate the value of a_μ, a_ν and ENCODE (a, v), and note the result as c, that is c = ENCODE (a, v) or $c = a_\mu, a_\nu$.
2. The search condition of **certain data querying** can be modified to
" a_code = c with $a_\mu = 1$ and $a_\nu = 0$ ".

The search condition of **possible data querying** can be modified to "a_code $\geq$ c $\wedge$ a_code and c = c" with $a_\mu \leq 1$ and $a_\nu \geq 0$.

4. Application: Knowledge towards Malaria

Malaria is the world's most important tropical disease and kills more people than any other disease. It has been an important public health problem in Tripura too. One of the important reasons behind this is their poor knowledge about malaria. In this study we collected the data of knowledge towards malaria of rural tribal communities of Tripura and implemented efficient IF rough query optimization technique on malaria knowledge via SQL server.

For the study, a pre-structured and pre-tested questionnaire was prepared on malaria knowledge. 216 persons were selected from the rural tribal communities of Tripura. A proforma was filled up with the required information after a verbal interview of the family members who were above 18 years of age. Before the interview a consent form was also signed by the subjects. The consent was in written form and was in an understandable language of the subjects.

The collected information is divided into the following tables (i) Socio-demographic characteristics (ii) Knowledge about malaria. We have used IF rough methodology to convert the qualitative or quantitative data into encoding, membership and non-membership values and implemented into IF rough algorithms. The encoding technique is shown as below:

Suppose cause of malaria is the multi-valued attribute about knowledge of the study population towards malaria. The domain and equivalence class of attribute "cause_malaria" are $D_{cause\_malaria}$ and $R_{cause\_malaria}$. These are defined as follows:

$\mathbf{D_{cause\_malaria}}$ = {Mosquito, Malaria parasite, Bad personal hygiene, Insanitary conditions, Contaminated water, God given, Man made, Evil spirit, Don't know}

$\mathbf{R_{cause\_malaria}}$ = {[Mosquito, Malaria parasite] [Bad personal hygiene, Insanitary conditions, Contaminated water] [God given, Man made, Evil spirit] [Don't know]}

Now, for instance to encode the arbitrary value (v) = {Mosquito, Contaminated water} of the multi-valued attribute "cause_malaria", it is compared with each equivalence class of $R_{cause\_malaria}$. Arbitrary value (v) exists for first two equivalence classes due to which the first two bits are 1 each, the remaining bits are 0. Therefore, the cause_malaria_code is 1100. (Appendix Table I)

Copyright ⓒ 2014 SERSC

4.1 Table Design

In the logical design phase, Data Definition Language (DDL) command is used to design tables (relations). CREATE command is used to define the relations. Brief descriptions of the tables are as follows:

SQL: CREATE TABLE STATEMENT

The **SQL CREATE TABLE statement** is used to create and define a table.

```
(1)   CREATE TABLE SOCIO-DEMOGRAPHIC INFORMATION
      (ID              DECIMAL(3),
      NAME             CHAR(50),
      SEX              CHAR(25),
      AGE              DECIMAL(3),
      COMMUNITY        CHAR(30),
      VILLAGE          CHAR(30),
      LITERACY         CHAR(30),
      MU               CHAR(10),
      NMU              CHAR(10),
      PRIMARY KEY (ID));
```

SOCIO-DEMOGRAPHIC INFORMATION: The socio-demographic table has attributes ID, name, sex, age, community, village, literacy, membership values (MU) and non-membership values (NMU).

```
(2)   CREATE TABLE KNOWLEDGE ABOUT MALARIA
      (ID                                DECIMAL(3),
      CAUSE OF MALARIA                   CHAR(145),
      CAUSE OF MALARIA_C                 CHAR(5),
      SYMPTOMS OF MALARIA                CHAR (145),
      SYMPTOMS OF MALARIA_C              CHAR (5),
      SEASON OF MALARIA                  CHAR (145),
      SEASON OF MALARIA_C                CHAR (5),
      SPREAD DISEASE                     CHAR (145),
      SPREAD DISEASE_C                   CHAR (5),
      SOURCE OF KNOWLEDGE                CHAR (145),
      SOURCE OF KNOWLEDGE_C              CHAR (5),
      PEOPLE COME TO KNOW ABOUT MALARIA CHAR(145),
      PEOPLE COME TO KNOW ABOUT MALARIA_C CHAR(5),
      FOREIGN KEY (ID));
```

KNOWLEDGE ABOUT MALARIA: This table shows knowledge of the study population towards malaria. The attributes are ID, cause of malaria, symptoms of malaria, commonest season of getting malaria, cause of spreading the disease, sources of knowledge regarding malaria and the methods by which the person come to know about malaria. Each attribute also contains the encoding of the respective attribute.

 Copyright ⓒ 2014 SERSC

SQL: INSERT STATEMENT

The **SQL INSERT statement** is used to insert the records into tables.

```
(1)     INSERT
              INTO SOCIO-DEMOGRAPHIC INFORMATION
        VALUES (1, Laljambul Halam, male, 61, Halam,
        Dugangpara, Illiterate, 0.8, 0.1);
 (2)    INSERT
              INTO KNOWLEDGE ABOUT MALARIA
        VALUES (1, mosquito, 1000, fever,10000, summer,
        01000, don't know, 00001, doctor, 01000, blood
        slide test, 01000);
```

Similarly, we can insert all values of the tuples in Tables.

5. SQL: IF ROUGH DATA QUERIES

Once the tables were created and populated with data, we evaluated their utility with sample queries. The queries are divided into two parts: certain data query and possible data query. Certain data querying is that search there objects fully matching the querying conditions and the querying results are obtained by the ***lower approximation*** of attributes values. Possible data querying is rough querying and the querying results are obtained by the ***upper approximation*** of attributes values. The queries and their results are as follows:

Question 1: Select ID of those respondents who are males, age ≥ 40, illiterate and know that mosquito as the cause of malaria followed by contaminated water.

Certain Data Query:

```
SELECT(Socio-demographic.ID),(Socio-demographic.age), (Socio-
demographic.sex),(Socio-demographic.Liteacy),(Knowledge About
Malaria.Cause malaria)
FROM Knowledge About Malaria, Socio-demographic
WHERE ((age ≥ 40) and (sex ='male')and (Literacy
='illiterate')and (MU = 1) and (NMU = 0)and (Cause of
malaria_c = '1100'))and (Socio-demographic.ID = Knowledge
About Malaria.ID);
```

Lower Approximation

Result [Table 1(a): Knowledge about Cause of Malaria]

ID	age	sex	Literacy	Causemalaria
3	75	male	Illiterate	mosquito,contaminatedwater
191	41	male	Illiterate	mosquito,contaminatedwater
207	70	male	Illiterate	mosquito,contaminatedwater

Copyright ⓒ 2014 SERSC

Possible Data Query:

```
SELECT (Socio-demographic.ID),(Socio-demographic .age)
,(Socio-demographic.sex),(Socio- demographic.
Literacy),(Knowledge About Malaria.Cause malaria)
FROM Knowledge About Malaria, Socio-demographic
WHERE ((age>=40) and (sex='male')and (Literacy
='illiterate')and (MU ≤ 1) and (NMU ≥ 0)and (Cause of
malaria_c ≥ '1100'))and ((Socio-demographic.ID = Knowledge
About Malaria.ID));
```

Upper Approximation
Results [Table 1(b): Knowledge about Cause of Malaria]

ID	age	sex	Literacy	Causemalaria
1	61	male	Illiterate	mosquito,contaminatedwater
1	62	male	Illiterate	mosquito,contaminatedwater
1	63	male	Illiterate	mosquito,contaminatedwater
3	75	male	Illiterate	mosquito,contaminatedwater
191	41	male	Illiterate	mosquito,contaminatedwater
207	70	male	Illiterate	mosquito,contaminatedwater

Explanation: From Table 1(a), we get the results of certain data querying are {3,191,207} and from Table 1(b), possible data querying are {1, 3,191,207}.

Question 2: Retrieve ID and name of those respondents who are Chakma and above 30 years old and mentioned that doctors as their source of knowledge about malaria.

Query:

```
SELECT (Socio-demographic .ID),( Socio-demographic
.Name),( Socio-demographic. Community), ( Socio-demographic.
Age),( Knowledge About Malaria .Source_of_Knowledge)
FROM Knowledge About Malaria, Socio-demographic
WHERE ((Community ='Chakma')and (Age > 30) and (MU = 1)and
(NMU = 0) and (Knowledge About Malaria. Source_of_knowledge_c
='01000'))and ((Socio-demographic.ID= Knowledge About
Malaria.ID));
```

 Copyright ⓒ 2014 SERSC

Result [Table 2: Doctors is the Source of Knowledge about Malaria]

ID	Name	Community	Age	Source_of_Knowledge
82	Gouranga Chakma	Chakma	41	Doctor
84	Halog Chakma	Chakma	50	Doctor
90	Sangita Chakma	Chakma	35	Doctor
92	Tara Chakma	Chakma	41	Doctor
95	Goraitien Chakma	Chakma	35	Doctor
97	Lalnghaka Chakma	Chakma	40	Doctor
107	Neisekbhum Chakma	Chakma	65	Doctor
136	Lalnghaka Chakma	Chakma	70	Doctor
137	Gouranga Chakma	Chakma	38	Doctor
142	Satya Chakma	Chakma	41	Doctor
148	Bideejoi Chakma	Chakma	50	Doctor
157	Neisekbhum Chakma	Chakma	70	Doctor
159	Goraitien Chakma	Chakma	35	Doctor
160	Haidan Chakma	Chakma	40	Doctor

6. Conclusion

This paper deals with the information on knowledge of the tribal people towards malaria of Tripura and presented a new efficient query optimization technique for the multi-valued or imprecise IF rough database on malaria knowledge. Then the querying is executed into SQL server to make the implementation of IF rough data querying simpler. Finally, we see that our solution makes the implementation of IF rough data querying much simpler and more efficient and can serve in a better way the purpose of medical experts.

References

[1] K. Atanassov, "Intuitionistic Fuzzy Sets, Fuzzy Sets and Systems", vol. 20, **(1986)**, pp. 87–96.
[2] T. Beaubouef and F. E. Petry, "Uncertainty modeling for database design using intuitionistic and rough set theory", Jour. Intelligent and Fuzzy Systems, vol. 20, no. 3, **(2009)**, pp. 105-117.
[3] E. F. Codd, "A relational model of data for large shared data banks", Comm. ACM, vol. 13, pp. 377–387, **(1970)**.
[4] C. Gangwal and R. N. Bhaumik, "Normalization of Intuitionistic fuzzy rough relational Databases", IJARCET, vol. 1, no. 6, **(2012)**, pp. 24-27.
[5] C. Gangwal and R. N. Bhaumik, "Intuitionistic fuzzy rough relational database model", International Journal of Database Theory and Application, vol. 5, no. 3, **(2012)**, pp. 91-102.
[6] C. Gangwal, R. N. Bhaumik and S. Kumar, "Application of IF rough relational model to deal with Diabetic patients, Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing", Lecture Notes in Computer Science (LNCS), vol. 8170, **(2013)**, pp. 191–199.
[7] S. Hiremath and P. Chandra, "Efficient approach for query optimization in rough data", Internat. J. of Sci and Research, vol. 2, no. 6, **(2013)**, pp. 239-242.
[8] X. Hu, X. Hong and Y. Yuan, "A high efficiency approach to querying rough data", Fuzzy system and knowledge discovery, vol. 2, **(2007)**, pp. 308-313.
[9] Z. Pawlak, "Rough sets", Internat. J. Comput. Inform. Sci, vol. 11, **(1982)**, pp. 341-356.
[10] L. A. Zadeh, "Fuzzy sets", Information Control, vol. l8, **(1965)**, pp. 338-353.

Copyright ⓒ 2014 SERSC

APPENDIX –

Table I - Encoding of Knowledge of the Study Population towards Malaria

Domain	Equivalence Class	Encoding
Cause of malaria		
Mosquito	[Mosquito, Malaria parasite]	1000
Malaria parasite		
Bad personal hygiene	[Bad personal hygiene, Insanitary conditions, Contaminated water]	0100
Insanitary conditions		
Contaminated water		
God given	[God given, Man made, Evil spirit]	0010
Man made		
Evil spirit		
Don't know	[Don't know]	0001
Symptoms of malaria		
Fever	[Fever, Chills, Rigors]	10000
Chills		
Rigors		
Headache	[Headache, Joint pains, Body aches]	01000
Joint pain		
Body ache		
Loss of Appetite	[Malaise, Nausea, Weakness, Loss of appetite, Vomiting]	00100
Vomiting		
Malaise		
Nausea		
Weakness		
Unconsciousness	[Unconsciousness]	00010
Don't know	[Don't know]	00001
Commonest season of getting malaria		
Winter	[Winter]	10000
Summer	[Summer]	01000
Rainy season	[Rainy season]	00100
Change of season	[Change of season, All seasons]	00010
All seasons		
Don't know	[Don't know]	00001
How does the disease spread		
Through mosquito	[Through mosquito]	10000
Houseflies	[Houseflies]	01000
Drinking contaminated water	[Drinking contaminated water, Eating contaminated food]	00100
Eating contaminated food		
Coughing	[Coughing, Sneezing]	00010
Sneezing		
Don't know	[Don't know]	00001
How does the person come to know whether he/she has malaria		
Symptoms	[Symptoms]	10000
Blood slide test	[Blood slide test]	01000
Urine test		

 Copyright ⓒ 2014 SERSC

Stool test	[Urine test , Stool test]	00100
X-ray	[X-ray]	00010
Don't know	[Don't know]	00001
Source of knowledge regarding malaria		
Family members	[Family members]	10000
Doctors	[Doctors, Health centre]	01000
Health centre		
Qualified private practitioner	[Qualified private practitioner, Health worker]	00100
Health worker		
Newspaper	[Newspaper, Television, Radio, Poster]	00010
Television		
Radio		
Poster		
Nothing	[Nothing]	00001

Copyright ⓒ 2014 SERSC

Copyright ⓒ 2014 SERSC