 ISSN: 2454-4248
121 – 125

Impute the Missing Data through Fuzzy Expert System for the Medical Data Diagnosis

B.S. Panda*
Assoc. Professor
MITS, Rayagada, Odisha.
bspanda.cse@gmail.com

Prasanta Khilla
M.Tech Student
MIPS, Rayagada, Odisha.
prasanta_3188@yahoo.co.in

Abstract: Data Processing with missing attribute values based on fuzzy sets theory. By matching attribute-value pairs among the same core or reduce of the original data set, the assigned value preserves the characteristics of the original data set. Malaria represents major public health problems in the tropics. The harmful effects of malaria parasites to the human body cannot be underestimated. In this paper, a fuzzy expert system for the management of malaria (FESMM) was presented for providing decision support platform to malaria researchers, The fuzzy expert system was designed based on clinical observations, medical diagnosis and the expert's knowledge. We selected 15 cases with Malaria and computed the missing results that were in the range of common attribute element by the domain experts.

Keywords: *Missing Data, Malaria, Fuzzy Logic, Knowledge base, Fuzzy Expert System.*

*****

1. Introduction

Missing attribute values commonly exist in real world data set. They may come from the data collecting process or redundant diagnose tests, unknown data and so on. Discarding all data containing the missing attribute values cannot fully preserve the characteristics of the original data. Various approaches on how to cope with the missing attribute values have been proposed in the past years.

Databases such as those which store measurement or medical data may become subject to missing values either in the data acquisition or data-storage process. Problems in a sensor, a break in the data transmission line or non-response to questions posed in a questionnaire are prime examples of how data can go missing. The problem of missing data poses a difficulty to the analysis and decision making processes which depend on this data, requiring methods of estimation which are accurate and efficient. Various techniques exist as a solution to this problem, ranging from data deletion to methods employing statistical and artificial intelligence techniques to impute for missing variables. However, some statistical methods, like mean substitution have a high likelihood of producing biased estimates1 or make assumptions about the data that may not be true, affecting the quality of decisions made based on this data.

2. Missing Data

Incomplete data are questions without answers or variables without observation. Even a small percentage of missing data can cause serious problems with the analysis leading to draw wrong conclusions and imperfect knowledge. There are many techniques to manipulate the imperfect knowledge and manage data with incomplete items, but no one is absolutely better than the others[9,10].

To handle such problems, researchers are trying to solve it in different directions and then proposed to handle the information system. The attribute values are important for information processing. In the field of databases, various efforts have been made for the improvement and enhance of database query process to handle the data. The different researchers have tried and are trying to handle the imprecise and/or uncertainty in databases. The methodology followed by different approaches like: Fuzzy sets, Rough sets, Boolean Logic, Possibility Theory, Statistically Similarity etc[14].

3. Fuzzy Logic

Fuzzy logic is usually used for building fuzzy rules that can be easily understood by humans. Therefore, it is common to describe fuzzy variables as linguistic variables. The linguistics variables that we will use in this research are mild, moderate, severe and very severe for both the input and output parameters in the fuzzy model. By using those linguistic variable, fuzzy if-then rules which are the main output of the fuzzy system would be set up: generally presented in the form of: if *x* is A then *y* is B where *x* and *y* are linguistic variables and A and B are linguistic values, determined by their fuzzy sets. The first part of the rule is called the antecedent, and can consist of multiple parts with the operators AND or OR between them. The latter part is called the consequent, and can also include several outputs.

Fuzzy logic [4,5] is determined as a set of mathematical principles for knowledge representation based on degrees of membership rather than on crisp membership of classical binary logic. This powerful tool to tackle imprecision and uncertainty was initially introduced by [1] to improved tractability, robustness and low-cost solutions for real world problems. Fuzzy sets have been applied in many fields in which uncertainty plays a key role. Medical diagnosis is an excellent example of vagueness and uncertainty. Fuzzy set theory is a response to the demand for ideas and approaches for handling nonstatistical uncertainty [2].

4. An Example of a Malaria Data Set with Missing Attribute values:

The attributes are *Temperature*, *Headache*, *Nausea*, *Vomiting*, *Joint Pain*, *Body Weakness*, *MP* and with the decision *Malaria*. However, many real-life data sets are incomplete, i.e., some attribute values are missing. In Table 1 missing attribute values are denoted by "?"s.

 ISSN: 2454-4248
121 – 125

Case	Temperature	Headache	Nausea	Vomiting	Joint Pain	Body Weakness	MP	Malaria
1	Mild	Mild	Mild	Mild	Mild	?	Mild	No
2	Moderate	Mild	Mild	?	Moderate	Moderate	Moderate	Yes
3	Severe	Moderate	Mild	Mild	Mild	Severe	Moderate	Yes
4	Very Severe	Mild	?	Mild	Severe	Severe	Severe	Yes
5	Moderate	Mild	Mild	Moderate	Moderate	?	Mild	No
6	Mild	Moderate	Moderate	Mild	?	Mild	?	No
7	Mild	Mild	Moderate	Moderate	Severe	Severe	Moderate	Yes
8	Moderate	?	?	Moderate	Moderate	Moderate	Moderate	Yes
9	Moderate	Mild	Moderate	Moderate	Moderate	Moderate	Moderate	Yes
10	Mild	Mild	Moderate	Moderate	?	Mild	?	No
11	?	Severe	Severe	Severe	Severe	Very Severe	Moderate	Yes
12	Moderate	Severe	Moderate	Severe	Moderate	Severe	Moderate	Yes
13	Mild	moderate	Moderate	Moderate	?	Moderate	Moderate	No
14	Severe	Severe	Moderate	Severe	Severe	Severe	Severe	Yes
15	Mild	Mild	Mild	Moderate	Mild	Severe	?	No

Table 1: Missing attributes in Malaria dataset

5. Fuzziffication

Fuzzification is the process of changing a real scalar value into a fuzzy value. This is achieved with different types of fuzzifiers. There are generally four types of fuzzifiers, which are used for the fuzzification process. They are: Trapezoidal fuzzifier, Triangular fuzzifiers, Singleton fuzzifier, and Gaussian fuzzifier [3]. Triangular fuzzifier which is widely used will be used in this research.

Fuzzification of data is carried out by selecting input parameters into the horizontal axis and projecting vertically to the upper boundary of membership function to determine the degree of membership.

The first step in the development of fuzzy logic based expert system is to construct fuzzy sets for the parameters. This is shown in equations (3) to (6) below. On the basis of domain experts' knowledge, both input and output parameters selected for this research were described with four linguistic variables (mild, moderate, severe and very severe). The range of fuzzy value for each linguistic is shown in table 1 below:

6. Fuzzy Expert System

The success of a Fuzzy Expert System depends[12,13] upon the opinion of the domain experts on various issues related to the study. The FES is a rule based system that uses fuzzy logic rather that Boolean logic [8]. It was developed based on the following key components:

- Knowledge Base component
- Fuzzification Component
- Inference Engine Component
- Defuzzification Component

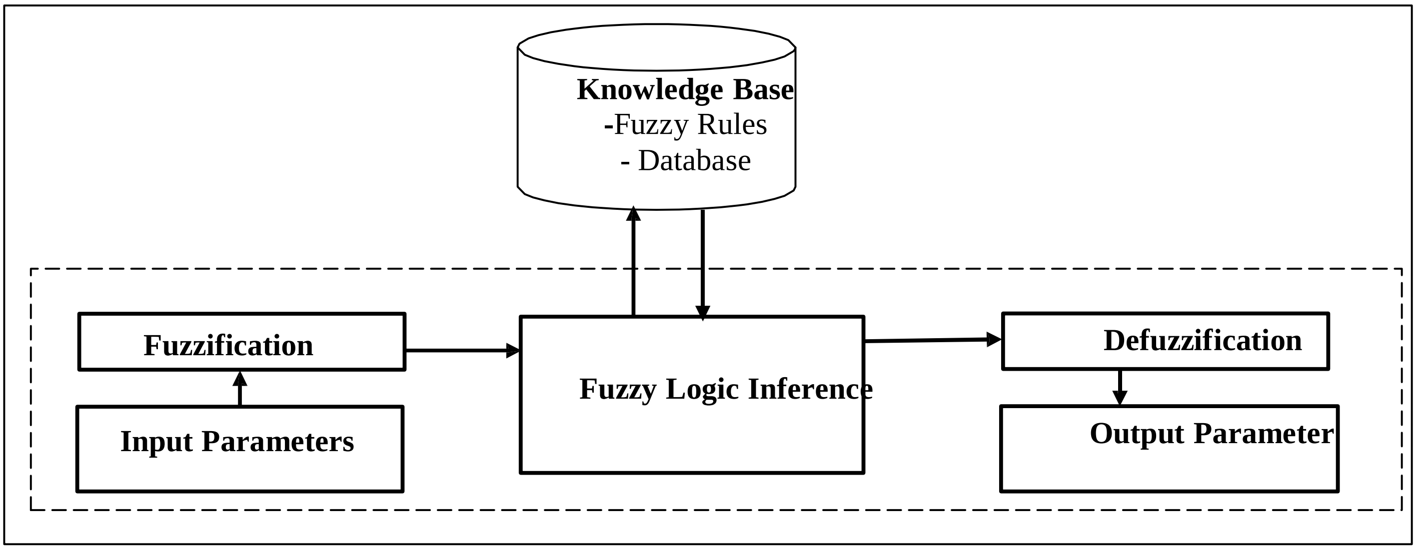

Figure 1: Architecture of FES

7. Fuzzy Inference Rules

The process of drawing conclusion from existing data is called inference. Fuzzy inference is the process of mapping from a given input to an output using the theory of fuzzy sets [3,4]. The core of decision making output is process by the inference engine using the rules contained in the rule base. The fuzzy inference engine uses the rules in the knowledge-base and derives conclusion base on the rules. FESMM inference engine uses a forward chaining mechanism to search the knowledge for the symptoms of a disease. For each rule, the inference mechanism looks up the membership values in the condition of the rule. Fuzzy inputs are mapped into their respective weighting factors and their associated linguistic variables to determine their degree of membership. The aggregation operator is used to calculate the degree of fulfillment or firing strength of a rule.

For this research, we have decided to apply fuzzy logical AND to evaluate the composite firing strength of the rules.

In practice, the fuzzy rules sets usually have several antecedents that are combined using fuzzy logical operators, such as AND, OR and NOT, though their definitions tend to vary: AND simply uses minimum weight of all the antecedents, while OR uses the maximum value. There is also the NOT operator that subtracts a membership function from 1 to give the “complementary” function. The IF part of a rule is called the “antecedent” and the THEN part is called the “consequent” [5].

Fuzzification begins with the transformation of the raw data using the functions that are expressed in equations (3) to (6) below. During the process, linguistic variables are evaluated using triangular membership function and are accompany by associated degree of membership ranging from 0 to 1 as shown in equations (3) to (6) below. These formulas are determined by aid of both the expert doctors in the field of tropical medicine and literature[11].

Some of the rules as follows:

Rule1: IF tempature = mild and headache = mild and nausea = mild and vomiting = mild and joint pain = mild and body weakness = ?and MP = mild THEN malaria = No.

Rule 2: IF tempature = moderate and headache = mild and nausea = mild and vomiting = ? and joint pain = moderate and body weakness = moderate MP = moderate THEN malaria = Yes.

Rule 3: IF tempature = severe and headache = moderate and nausea = mild and vomiting = mild and joint pain = mild and body weakness = severe MP = moderate THEN malaria = Yes.

Rule 5: IF tempature = moderate and headache = mild and nausea = mild and vomiting = modarate and joint pain = modarate and body weakness = ? MP = mild THEN malaria = No.

Rule 8: IF tempature = moderate and headache = ? and nausea = ? and vomiting = modarate and joint pain = modarate and body weakness = modrate MP = modrate THEN malaria = Yes.

Rule 10: IF tempature = mild and headache = mild and nausea = modarate and vomiting = modarate and joint pain = ? and body weakness = mld MP = ? THEN malaria = No.

Rule 13: IF tempature = mild and headache = modarate and nausea = modarate and vomiting = modarate and joint pain = ? and body weakness = modrate MP = mild THEN malaria = No.

Rule 15: IF tempature = mild and headache = mild and nausea = mild and vomiting = modarate and joint pain = mild and body weakness = severe MP = ? THEN malaria = Yes.

A rule is said to fire if any of the precedence parameters (mild, moderate, severe, and very severe) evaluate to true/yes (1); other, if all the parameters evaluate to false/no (0), it does not fire.

8. Fuzzy Membership Function

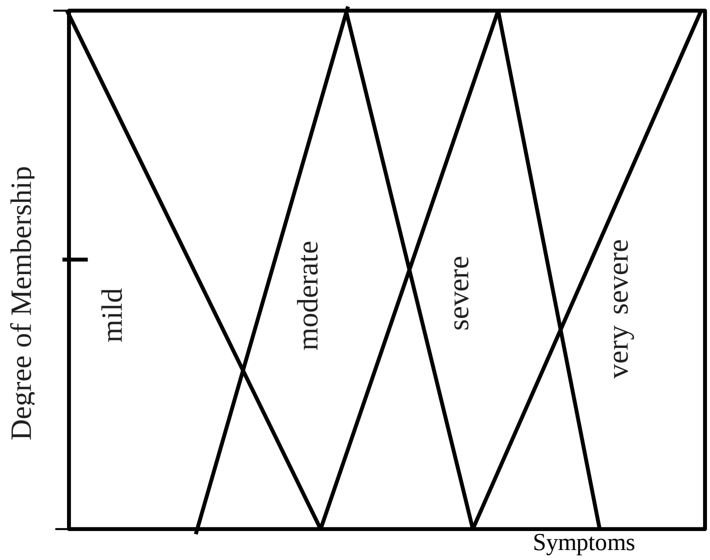

Fig. 2. Fuzzy Membership Function

9. The Algorithm For Fuzzy Diagnosis

The developed algorithm for the fuzzy diagnostic process of malaria is:

Step 1 : Input signs and symptoms of patient complaint into the system. Where **m** = number of signs and symptoms.
Step 2: Search the Missing Data **d,** which has the signs and symptoms identified.
Step 3: Get the weighing factors (wf) (the associated degree of intensity) wf = 1, 2, 3, 4 Where 1 = Mild, 2 = Moderate, 3 = Severe, 4 = Very Severe.
Step 4: Apply fuzzy rules.
Step 5: Map fuzzy inputs into their respective weighing factors to determine their degree of membership.
Step 6: Determine the rule base evaluating (non-minimum values).
Step 7: Determine the firing strength of the rules R.
Step 8: Calculate the degree of truth R, of each rules by evaluating the nonzero minimum value.
Step 9: Compute the intensity of the disease.
Step 10: Output fuzzy diagnosis.

10. Defuzzification

The defuzzification process translates the output from the inference rule into crisp output. This is due to the fact that, the output from the inference engine is usually a fuzzy set while for most medical applications, crisp values are required. The input to the defuzzification process is a fuzzy set while the output of the defuzzification process is a single number (crisp output). Many defuzzification techniques are proposed and four common defuzzification techniques are: center-of-area (gravity), center-of-sums, max-criterion and mean of maxima.

11. Research Experiment

In this method, one of the simplest methods to handle missing attribute values, such values are replaced by the most common value of the attribute. In different words, a missing attribute value is replaced by the most probable known attribute value, where such probabilities are represented by relative frequencies of corresponding attribute values.

We considered a set of five diseases **D**, and the expert doctors defined a set of signs and symptoms **M** relevant to a particular tropical disease.

D = {d_1, d_2, d_3, d_4, d_5} where d_1, d_2, d_3, d_4, d_5 represents the five tropical diseases under consideration.

M = {m_1, m_2, m_3, m_4 . . . m_n} where m_1, m_2, m_3, m_4 . . . m_n represents the signs and symptoms of a particular tropical disease.
To specify the signs/symptoms intensity for a particular patient, the expert doctors applied the common symptoms to the missing symptoms and weighing factors to the set **M**, thereby assigning fuzzy values to the signs/symptoms. The fuzzy values are selected from the fuzzy set:

{Mild (1), Moderate (2), Severe (3), Very Severe (4)}

Patients' state of health (with respect to malaria) was evaluated by the domain expert based on signs, symptoms and investigations. The intensity of signs, symptom and investigation was rated as mild (1), moderate (2), severe (3), and very severe (4). Table 3 below shows the weights assigned to patients after an interactive session with the expert doctors.

12. Research and Discussion

We have made humble attempt to implement the concept of Fuzzy Rule Based Systems that incorporated fuzzy techniques to missing the attributes and simplifying the diagnosis of malaria. A fuzzy expert system for diagnosis malaria was developed. In the fuzzy logic implementation, the selection of fuzzifier, rule base and inference engine determined the output of FESMM. We choose triangular fuzzifier, the rule base was designed based on knowledge of domain experts (five medical doctors), and the inference technique we employed was RSS. Fuzzy logic was utilized to remove uncertainty, ambiguity and vagueness inherent in medical diagnosis.

13. Fuzzy Results

Case	Common Attribute	Diagnosis (Result)
001	1- 6	No
002	2-4	Yes
003	0	Yes
004	3-3	Yes
005	2-3	No
006	2-3	No
007	0	Yes
008	2-5	Yes
009	0	Yes
010	78	No
011	3-4	Yes
012	0	Yes
013	2-5	No
014	0	Yes
015	1-4	No

14. Conclusions

On the basis of the all presented, it can be concluded that there is no doubt whether Fuzzy Expert Systems should be applied for medical purpose. The use of fuzzy logic for medical diagnosis provides an efficient way to assist inexperienced physicians to arrive at the final diagnosis of malaria more quickly and efficiently. The developed

FESMM provides decision support platform to assist malaria researchers, physicians and other health practitioners in malaria endemic regions. Furthermore, implementation of FESMM will reduce doctors' workload during consultation and ease other problems associated with hospital consultations.

References

[1] L.A. Zadeh, Fuzzy sets and systems: In fox, J. (ed.): Proceedings Symposium on System Theory, Polytechnic Institute of Brooklyn, (1965), 29-37.

[2] Grzymala-Busse J.W. Knowledge acquisition under uncertainty - A rough set approach. Journal of Intelligent & Robotic Systems. (1988) 3-16.

[3] L.H. Tsoukalas, and R.E. Uhrig. Fuzzy and neural approaches in Engineering, John Wiley & Son, Inc. (1993).

[4] H.P. Nguyen and V. Kreinovich, Fuzzy logic and its applications in medicine. International Journal of Medical Informatics, 62(2001), 165-173.

[5] H.T. Nguyen and E.A. Walker, A First Course in Fuzzy Logic, Application in Intelligent System, Boston; Kluwer Academic, 1997.

[6] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, A multilayer perception-based medical decision support system for heart disease diagnosis, Expert Systems with Applications, 30(2006), 272-281.

[7] S.F. Tohal, and U.K. Ngah, Computer aided medical diagnosis for the identification of malaria parasites, IEEE -ICSCN, MIT Campus, Anna University, Chennai, India, (2007), 521-522.

[8] W.W. Melek, and A. Sadeghian, A theoretical framework for intelligent expert systems in medical encounter evaluation. Expert Systems, 26(2009), 82-99.

[9] Grzymala-Busse J.W. Data with missing attribute values: Generalization of indiscernibility relation and rule induction. Transactions on Rough Sets, Lecture Notes in Computer Science Journal Subline, Springer-Verlag, vol. 1 (2004) 78-95.

[10] Lipski W. Jr. On semantic issues connected with incomplete information databases. ACM Transactions on Database Systems 4 (1979), 262-296.

[11] S.F. Tohal, and U.K. Ngah, Computer aided medical diagnosis for the identification of malaria parasites, IEEE -ICSCN, MIT Campus, Anna University, Chennai, India, (2007), 521-522.

[12] W.W. Melek, and A. Sadeghian, A theoretical framework for intelligent expert systems in medical encounter evaluation. Expert Systems, 26(2009), 82-99.

[13] B. S. Panda, Rahul Abhishek, Gantayat, S. S. "Uncertainity Classification of Expert Systems - A Rough Set Approach" in: (ISCON – 2012) International Journal of Computer Applications ISBN: 973-93-80867-87-0, 2012.

[14] S. S. Gantayat, Ashok Misra, B. S. Panda, "A Study of Incomplete Data – A Review" In: FICTA-2013, LNCS Springer. Pp 401-408. ISBN: 978-3-319-02930-6.

[15] B. S. Panda, S. S. Gantayat, Ashok Misra (2017) " Approach and Impute the Missing Data using Rough Set Rule Induction Method" In: International Journal Of Engineering, Science And Mathematics (IJESM), ISSN : 2320-0294 Vol. 6, Issue 8.

.