Performances of Estimating Null Values using Noble Evolutionary Algorithm (NEAs) by Generating Weighted Fuzzy Rules

Muhammad Firoz Mridha Department of Computer Science Stamford University Bangladesh Dhaka, Bangladesh

ABSTRACT

This paper Present a noble technique to estimate null values from relational database systems. At present some methods exist to estimate null values from relational database systems. The estimated accuracy of the existing methods are not good enough. We have used an advance technique for estimating null values in relational database systems. In our paper we present the technique to generate weighted Fuzzy rules from relational database systems for estimating null values using Noble Evolutionary algorithms. The parameters (operators) of the Evolutionary algorithms are adapted via Fuzzy systems. We have fuzzified the attribute values using membership functions shape. The results of the evolutionary algorithms are the weights of the attributes. The different weights of attribute generate a set of Fuzzy rules. From this we have obtained a set of rules. Our proposed techniques have a higher average estimated accuracy rate and able to estimate the null values in relational database systems.

Keywords

Fuzzy System, Membership Functions, Noble Evolutionary algorithms, Null values, Relational Database Systems, Weighted Fuzzy Rules.

1. INTRODUCTION

Fuzzy systems have become popular components of consumer products because they are able to solve difficult nonlinear control problem, exhibit robust behavior and present linguistic representations. These rule-based systems are more suitable for complex system problems where it is very difficult to describe the system mathematically. One of the most important considerations in designing any fuzzy system is the generation of the fuzzy rules as well as membership functions for each fuzzy set. This paper present NEAs approach to solve problem. The solving procedure mainly based on Evolutionary algorithms. It has been observed that there are many drawbacks in the early methods in estimating null values. We have estimated null values more accurately as well as to overcome the drawbacks of the previous methods. Global optimization problems are very difficult to solve. In order to understand the difficulties it is important to note that all local optimization techniques can at most locate a local minimum.

2. FUZZY EXPERT SYSTEM

The fuzzy expert system works as follows [1]: 1).Determine the fuzzy membership values activated by the inputs.2). Determine which rules are fired in the rule set. 3). Combine the membership values for each activated rule using the AND operator.4). Trace rule activation membership values back through the appropriate

Manoj Banik Department of Computer Science and Engineering Ahsanullah University of Science and Technology, Dhaka, Bangladesh

output fuzzy membership functions.5). Utilize defuzzification to determine the value for each output variable. 6). Make decision according to the output values.

2.1 Membership Functions

A membership function is a curve that defines how each point in the input space is mapped to a membership value as shown in Fig.1 and Fig 2.





Membership



Fig.2 Membership function for Experience attributes

In this paper, the membership functions of the linguistic term, "L", "SL", "M", "SH", and "H" of the attributes and "Experience" in the relational database system are adopted as shown in Fig. 2[3].

2.2 Fuzzy Rule Base

The general form of a fuzzy rule in a fuzzy system is [1] If x₁₁ is S1, and x 2 is S2, ..., x k is Sk Then y₁₁ is T1, ..., and y1 is T1

2.3. Weighted Fuzzy Rules

Weighted fuzzy rules are a set of rules including the weights of the attributes; w_{ij} denotes the weight of attributes A_i of the ith rule in the rule

RULE BASE CONTAINS WEIGHTED FUZZY RULES

Rule m: IF $A_1=a_{m1}(W=w_{m1})$ AND $A_2=a_{m2}(W=w_{m2})$ ANDAND $A_n=a_{mn}(W=w_{mn})$ THEN $B=t_m$

2.4 BASIC CONCEPT OF FUZZY SETS:

In a fuzzy set, each element of the set is associated with a membership value between 0 and 1 described by a membership function to indicate the grade of membership of the element in the fuzzy set. There are two types of membership to represent fuzzy sets. One is the discrete type membership function, and the other is the continuous type membership function. Let U be the universe of discourse, $U = \{u_1, u_2, ..., u_n\}$. A fuzzy set universe of discourse U can be represented as follows:

$$A = \mu_A(u_1)/u_{1+}\mu_A(u_2)/u_{2+...+}\mu_A(u_n)/u_n$$

Where μ_A is the membership function of the fuzzy subset A, $\mu_A: U \rightarrow [0,1]$, and $\mu_A(u_i)$ indicate the grade of membership of u_i in fuzzy subset A. If the universe of discourse U is a continuous set then the fuzzy subset A can be represented as follows: $A = /_u \mu_A(u)/u, \quad u \in U$

A linguistic term can be represented by a fuzzy set represented by a membership function.

In our paper, the membership functions of the linguistic term, "L", "SL", "M", "SH", and "H" of the attributes "Salary" and "Experience" in the relational database system are adopted as shown in Fig.1 and Fig. 2 respectively, where "L" denotes "Low", "SL" denotes "Somewhat Low", "M" denotes "Medium", "SH" denotes "Somewhat High" and "H" denotes "High".

3. NULL VALUES

NULL is special value for representing data we don't have. But null value is not a constant. In relational database system the attribute that has no values is defined as null values.

4. NONLE EVOLUTIONARY ALGORITHM

Evolutionary algorithms are specified for parameter optimization problems [5]. We used Evolutionary algorithms to generate weighted fuzzy rules.

4.1 CROSSOVER

In the crossover the SBMAC is used to produce the offspring population. For each subpopulation, μ offspring are generated.

4.2 MUTATION

In the mutation phase, the TVM operator is used to mutate all variables of an offspring [5].

4.3 Pseudo-code structure of NOBLE EVOLUTIONARY Algorithm

The general pseudo-code type structure of the noble evolutionary algorithm is shown below[5].

Noble_Evolutionary_Algorithm ()

{

}

t =0; / * Initialization the generation counter * / Initialize _Populatio n(); Evaluate_P opulation(); while(NOT termination condition satisfied) do { Apply_SBMA C(); / * Crossover operation * / Apply_TVM(); / * Mutation operation * / Evaluate_Population(); Alternate_ Generation (); t; /* Increase the generation counter * / }

5. RELATED WORKS AND VOLUTIONARY FUZZY SYSTEM

In this section, we review the Chen–and-Chen's and Chen-and-yeh's method for estimating null values in relational database systems [3]. First, we can use fuzzy similarity matrices to represent relations and null value can be estimated by the closeness degree of the tuple with respect to the closest rule and use expert knowledge base [3]. Fuzzy systems generally use expert knowledge base. But in complex or simple environment fuzzy systems are not designed efficiently by using expert knowledge. Moreover, for better system performance, it is more difficult and time consuming for experts to define a complete rule sets for complex system problems, which use large number of parameters. In this paper, in order to overcome the drawback of [3], we improve the methods presented in [3] and proposed a new method to estimate null values in relational database systems using Noble Evolutionary algorithm (NEAs).

6. METHODOLOGY

The implementation of the weighted fuzzy system is written in c++ and complied using the Borland c++ 4.5 compilers.

6.1 Attributes Weight Calculation

We represent a method to calculate the weight of attributes using NEAs for estimating null values in relational database systems.

7. FORMAT OF A CHROMOSOME

Let us consider a relation of a relational database shown in Table 2 Based on Fig 1 and Fig 2, the values of attributes "Degree" and "Experience" shown in Table I can be fuzzified into in Table II. First, we find the format of chromosome as shown in Fig3. where the value of each gene in chromosome as a real value between zero and one and 13th gene labeled B-L shown in Fig. 4 denotes the fuzzified values of the attributes "Degree" and "Experience" are "Bachelor" (B) and "Low" (L) respectively (e.g.. the tuples whose EMP-ID are S6 and S17 as shown in Table3.

Generation verses Real Value																	
Ge ne	no. +	1		2	3	4	5	5 6	i 7	8	9	10	11	12	13	14	15
		Real	Re	al	Real	Rea	l Re	al Re	al Rea	al Real	. Real	Real	Real	Real	Real	Real	Real
	L	Value	e Va	hue	Valu	e Val	ue Va	due Va	due vab	ue Valu	e Valu	e Value	e Value	value	value	Value	Value
	B-H M-H P-H B-SH M-SH P-SH B-M M-M P-M B-SL M-SL P-SL B-L M-L P-L Fig: 3: Format of a chromosome																
<u>Gene</u>	.no	, 1		2	3	4	5	6	7	8	9	10	11	12	13	14	15
		0.017	0.06	9	0.386	0 <i>5</i> 43	0.505	0.404	0.495	0.089	0.667	0.778	0.404	0.858	0.858	0.687	0.435
	F	8-H	M-H]	P-H	B-SH	M-SH	P-SH	B-M	M-M	P-M	B-SL	M-SL	P-SL	B-L	M-L	P-L
Fight Example of a sharmona and																	

1 7 7 1

...

Fig4: Example of a chromosome

TABLE I: Relational database with Null values

EMP-	DEGREE	EXPERIENCE	SALARY
ID			
E1	Ph.D.	7.2	63000
E2	Master	2.0	37000
E3	Bachelor	7.0	40000
E4	Ph.D.	1.2	47000
EΣ	Master	7.5	53000
E6	Bachelor	1.5	26000
E7	Bachelor	2.3	29000
E8	Ph.D.	2.0	50000
E9	Ph.D	3.8	54000
E10	Bachelor	3.5	35000
E11	Master	3.5	40000
E12	Master	3.5	41000
E13	Master	10.0	68000
E14	Ph.D.	5.0	57000
E15	Bachelor	5.0	36000
E16	Master	6.2	50000
E17	Bachelor	0.5	23000
E18	Master	7.2	55000
E19	Master	6.5	51000
E20	Ph.D.	7.8	65000
E21	Master	8.1	64000
E22	Ph.D	8.5	NULL

From Fig.3 we can see that each chromosome represents a combination of the weights of attributes, and it is a string of the attributes which will be used to estimate null values in relational database systems. A population contains a set of chromosomes, and we can arbitrary set the number of chromosomes in a population. In this chapter, we let a chromosome consist of 15 genes. Because the total weights of the attributes must be equal to one, the weight of attribute "Experience" must equal to one minus the weight of attribute "Degree". For example, we assume that there is a chromosome as shown in Fig4. Assume that we want to estimate the null value of the attribute "Salary" of a tuple whose fuzzified value of the attributes are "Degree" and "Experience" are "Ph.D." (P) And "Some what low" (SL) then from Fig.4 we can see that the values of the gene labeled

(P-SL) is 0.858 and 0.131 (i.e., 1-0.858=0.142), respectively, to calculate the degrees of closeness between the tuple which contains the null value and other tuples in the database, respectively. Therefore, the contents of the chromosome shown in Fig. 5 can be translated n the following 15 rules.

TABLE II: Fuzzified value of Degree and Experience

EMP-	DEGREE	EXPERIENCE	SALARY
ID			
El	Ph.DA.O	09	63000
E2	Master/1.0	05	37000
E3	Bachebr/1.0	1.0	40000
E4	Ph.DA.O	09	47000
ES	Master/1.0	0.75	53000
E6	Bachebr/1.0	0.75	26000
E7	Bachebr/1.0	0.65	29000
E8	Ph.DA.O	05	50000
E9	Ph.DA.O	0.6	54000
E10	Bachebr/1.0	0.75	35000
E11	Master/1.0	0.75	40000
E12	Master/1.0	0.7	41000
E13	Master/1.0	1.0	68000
E14	Ph.D/I.0	1.0	57000
E15	Bachebr/1.0	1.0	36000
E16	Master/1.0	0.6	50000
E17	Bachebr/1.0	1.0	23000
E18	Master/1.0	09	55000
E19	Master/1.0	0.75	51000
E20	Ph.DA.O	0.6	65000
E21	Master/1.0	0.55	64000
E22	Ph.D/I.O	0.75	70000

Rule 1: If Degree=Bachelor AND Experience=High, THEN the weight of degree=0.017 AND the weight of Experience=0.983.

Rule 2: If Degree=Master AND Experience=High, THEN the weight of degree=0.069 AND the weight of Experience=0.931.

Rule 3: If Degree=Ph.D. AND Experience=High, THEN the weight of degree=0.386 AND the weight of Experience=0.614.

Rule 4: If Degree=Bachelor AND Experience=Somewhat High, THEN the weight of degree=0.543 AND the weight of Experience=0.457.

Rule 5: If Degree=Master AND Experience=Somewhat High, THEN the weight of degree=0.505 AND the weight of Experience=0.495.

Rule 6: If Degree=Ph.D. AND Experience=Somewhat High, THEN the weight of degree=0.404 AND the weight of Experience=0.596

Rule 7: If Degree=Bachelor AND Experience=Medium, THEN the weight of degree=0.495 AND the weight of Experience=0.505.

Rule 8: If Degree=Master AND Experience=Medium, THEN the weight of degree=0.089 AND the weight of Experience=0.911.

Rule 9: If Degree=Ph.D. AND Experience=Medium, THEN the weight of degree=0.667 AND the weight of Experience=0.333.

Rule 10: If Degree=Bachelor AND Experience=Somewhat Low, THEN the weight of degree=0.778 AND the weight of Experience=0.222.

Rule 11: If Degree=Master AND Experience=Somewhat Low, THEN the weight of degree=0.404 AND the weight of Experience=0.596.

Rule 12: If Degree=Ph.D. AND Experience=Somewhat Low, THEN the weight of degree=0.858 AND the weight of Experience=0.142.

Rule 13: If Degree=Bachelor AND Experience=Low, THEN the weight of degree=0.858 AND the weight of Experience=0.142.

Rule 14: If Degree=Master AND Experience=Low, THEN the weight of degree=0.687 AND the weight of Experience=0.313.

Rule 15: If Degree=Ph.D. AND Experience=Low, THEN the weight of degree=0.435 AND the weight of Experience=0.565.

8. SUMMARY OF OUR METHOD

8.1 RULE BASE:

A rule base is used to indicate relationships in which some attributes determines other attributes. For example, Table I shows a set of rules including the weights of the attributes, where all rules in the rule base are given by experts, w_{ij} denotes the weight of attributes A_j of the ith rule in the rule base, $w \in [0, 1], 1 \le i \le n$, and $1 \le j \le n$.

8.2 ESTIMATING NULL VALUES:

The basic idea of our method is rule base shown in the Table I is closed to the tuple having a null value. The null value can be estimated by the closeness degree of the tuples with respect to the closest rule. Assume that, there is a relation in a relational database system having attributes A_1 , A_2 ,..., A_n and B, and assume that the attributes A_1 , A_2 ,..., and A_n determine the attribute B. Let " r_j , A_k " denote the value of attribute A_k appearing in the antecedent portion of rule r_j and let " T_i . A_k " denote the value of attribute B, then the value can be estimated as follows.

If the attribute B is defined in a numerical domain, and there rule r_i , where $1 \le i \le m$, in the base shown as follows:

IF
$$A_1{=}a_{j1}(W{=}w_{j1})$$
 AND $A_2{=}a_{j2}(W{=}w_{j2})$ ANDAND $A_n{=}a_{jn}(W{=}w_{jn})$ THEN $B{=}N_j$

8.3 CALCULATION OF THE FITNESS FUNCTION

The fitness function measures the performance of the system. In the following, we present a method to calculate the degree of closeness between two tuples. The ranks of the terms:

> Rank (Bachelor)=10, Rank (Master)=20, Rank (Ph.D)=30.

Suppose X be a nonnumeric attribute. The degree of the closeness Closeness (T_i,T_j) between two tuples T_i and T_j can be calculated by the following rules

If $Rank(T_i.X) >= Rank(T_j.X)$ then

$$\label{eq:closeness} \begin{split} Closeness(T_i,T_j) &= Similarity(T_i.X,T_j.X) \times Weight(T_j.Degree) + \\ & \frac{Ti.Experie\ nce}{Tj.Experie\ nc} \times Weight(T_j.Experience) \end{split}$$

Where Similarity $(T_i.X,T_j.X)$ denotes the degree of similarity between $(T_i.X)$ and $(T_j.X)$, and its value is obtained from a fuzzy similarity matrix of the linguistic terms of the attribute X defined by a domain expert.

Suppose that T_i , T_j , T_k be three tuples in a relational database. Assume that the degree of closeness between tuple T_i and T_j is denoted as Closeness (T_i,T_j) and the degree of closeness between tuple T_i and T_k is denoted as Closeness (T_i,T_k) After calculating the degree of closeness of the other tuples in the database with respect to T_i , the system will pick a tuple which is closet to tuple T_i , then we can calculate the estimated value "E.T_i. Salary " of the attribute "Salary" of the tuple T_i as follows:

E.T_i.Salary = T_i .Salary × Closeness (T_i, T_i)

Where T_i . Salary denotes the value of the attribute "Salary" of the tuple T_i .

We can calculate the Errors in the following way

$$Error = \frac{E.Ti .Salary - Ti .Salary}{Ti .Salary .}$$

Now average estimated Error of the tuples based on the combination of weights of the attributes derived from the chromosome, where

Avg Error=
$$(\sum_{i=1}^{n} Error_i)/n$$
.

Then we can obtain the fitness degree of this chromosome as follows:

Fitness Degree =1-Avg_Error.

8.4 Encoding Method

Fig.5 and Fig.6 shows the NES coding in detail. Each individual consists of two parts. This involves the Gaussians of all antecedents [1].



Fig 6: Fuzzy model encoding

8.5 CROSSOVER OPERATION

In our research we set the crossover rate α to 1.0. Therefore after the sections operations, the number of chromosomes in a population will continue to perform the crossover operations, where the system randomly picks up two chromosomes as the parents and randomly picks a crossover point. Then, the system performs the crossover operations on these two chromosomes at this crossover point to generate their two children.

8.7 FUZZY RULE

Fuzzy membership function can have different shapes. Rules obtained by simulation for Null values in relational database given Figure 7(a, b, c).

Rule 1:



Figure 7 (b): Triangular Curve



Figure 7(c): Triangular Curve

8.8 ESTIMATED NULL VALUES IN RELATIONAL DATABASE

In order to estimate null values in the attribute "Salary" of the tuple $T_{22 \text{ whose}}$ emp-id is S_{22} we must find a tuple which is closest to the tuple T_{22} . The process for computing the degree of closeness between two tuple is illustrated as follows. Let us consider the tuple T_1 shown in the table whose emp id is S1. Both degree is Ph.D that is similar. i.e similarity(T_{22} .Degree, T_1 .Degree)=1. Then based on the values of the attribute "Experience" of the tuple T_1 and the tuple T_{22} , we can get

$$\frac{\text{T22 E xperience}}{\text{T1 Experience}} = \frac{8.5}{7.2} = 1.180$$

We see that the fuzzified value of the attribute experience of the tuple T1 whose emp id is S1 is SH. Therefore we pick the value of the 6^{th} gene label "P-SH" is 0.303. It means that the weight of attribute Degree and Experience are 0.303 and .693 respectively. The Degree of Closeness between the tuple T22 and T1 can be calculated as follows From closeness rule we get

Closeness (T_{22} , T_1) = 1×0.268 + 1.1180× (1-0.268) = 0.303+0.822 = 1.125

After the degrees of closeness of all other tuples with respect to the tuple T_{22} are calculated we can see the tuple T_{20} is closest to the tuple the tuple T_{22} , where the degree of closeness Closeness (T_{22}, T_{20}) Between the tuples T_{22} and T20 is calculated as follows:

closeness(
$$T_{22}$$
, T_{20}) = 1 * 0.224 + $\frac{8.5}{7.2}$ * (1 - 0.224)
=0.224+1.0897*.776
=1.0698

Now the salary of tuple T_{22} can be calculated as ,follows $\begin{array}{rl} ET_{22} &=& 65000 \times 1.0698 \\ &=& 69065.83 \end{array}$ Error of calculation

$$\operatorname{Error}_{22} = \frac{69540 - 70000}{70000}$$
$$= -0.006$$

Comparison of the estimated result of the proposed method with the existing method									
				Chen-and-Chen's		Chen-and	-Yeh's	Our Proposed	
				Method[3]		Method[4]		Method	
EMP_ID	Degree	Experience	Salary	Salary Estimated	Estimated Error	Salary Estimated	Estimated Error	Salary Estimated	Estimated Error
1	Ph.D.	7.2	63000	63000	+0.000	65000	+0.032	62408	-0.009
2	Master	2	37000	33711	-0.089	30704	-0.170	38420	+0.038
3	Bachelor	7	40000	46648	+0.166	35000	-0.125	43200	-0.080
4	Ph.D.	1.2	47000	36216	-0.229	46000	-0.021	46029	-0.020
5	Master	7.5	53000	56200	+0.060	54500	+0.028	54125	+0.021
6	Bachelor	1.5	26000	27179	+0.045	26346	+0.013	27125	+0.004
7	Bachelor	2.3	29000	29195	+0.007	28500	-0.017	32000	+0.103
8	Ph.D.	2	50000	39861	-0.203	50000	+0.000	49845	-0.003
9	Ph.D.	3.8	54000	48061	-0.10	55000	+0.019	55042	+0.019
10	Bachelor	3.5	35000	32219	-0.079	31538	-0.099	36565	+0.044
11	Master	3.5	40000	40544	+0.014	41590	+0.040	40205	+0.005
12	Master	3.6	41000	41000	+0.000	45159	+0.101	40581	-0.010
13	Master	10	68000	64533	-0.051	65000	-0.044	67756	-0.003
14	Ph.D.	5	57000	55666	-0.023	55000	-0.035	62525	+0.096
15	Bachelor	5	36000	35999	-0.000	35000	-0.028	35241	-0.021
16	Master	6.2	50000	51866	+0.037	48600	-0.028	50457	+0.009
17	Bachelor	0.5	23000	24659	+0.072	25000	+0.087	22536	-0.020
18	Master	7.2	55000	55200	+0.004	52400	-0.047	53764	-0.026
19	Master	6.5	51000	52866	+0.037	49500	-0.029	51209	+0.004
20	Ph.D.	7.8	65000	65000	+0.000	65000	+0.000	65625	+0.009
21	Master	8.1	64000	58200	-0.091	58700	-0.083	62372	-0.025
22	22 Ph.D. 8.5		70000	67333	-0.038	65000	-0.071	69540	-0.006
Average	estimated E	rror			+0.062		+0.051		+0.006

TABLE III: Comparison

6. CONCLUSION

In this paper Based on the best chromosome we have estimated the Null values of attributes "salary" where the attribute contain the different weight of degree and experience .After a predefined number of evolution of the NEA the best chromosome contain the optimal weight of the attribute and they are translated into a set of rules to be used for estimating Null values. This proposed method can get a higher average estimated accuracy rate.

REFERENCES

- Yuhui Shi, Russell Eberhart, Yaobin Chen, "Implementation of Evolutionary Fuzzy Systems", IEEE Transaction of Evolutionary Computation, Vol-7, No.2, April 1999.
- [2] S.M.Chen and M.S.yeh."Generating fuzzy rules from relational database systems for estimating null values."Cybern.Syst.vol.28,no.8.pp.665-723,1997.
- [3] S.M.chen and H.H. Chen ,"estimating Null values in the distributed relational databases environment." Cybern. Syst.,vol.31.no.8,pp,851-871,2000.
- [4] H.L.Lin and S.M. Chen, "Generating weighted Fuzy rules from training data for handling fuzzy classification problems," in

- Proc.2000 Int. Computer Symp.: workshop Artificial Intelligence, Chiayi, Taiwan ,R.O.C.,2000, pp.11-18.
- [5] M.M.A Hashem ,Keigo Watanabe "evolutionary Computations", Khulna University of Engineering and

Technology, Bangladesh

- [6] J.C.Cubero and M.A. Vila,"A new definition of fuzzy functional dependency in fuzzy relational databases," Int.J.intel.Syst., vol.9.no.5, pp.441-448, 1994.
- [7] J.grant,"Null values in a relational database," inform.Processing Lett., vol.6,no.5,pp.156-157,1977.
- [8]C.M.Haung and S.M.Chen,"A new method to estimate null values in relational database systems using geneticalgorithms,"inProc.6thConf.ArtificialIntelligenceAppl ications,Kaushiung,Taiwan,R.O.C.,2001,2000,pp.599-604.
- [9] D.G.Burkhardt and P.P.Bonissone."Automated fuzzy knowledge base Generation and tuning." In Proc.1992 IEEE Int.Conf. Fuzzy Systems San Diego, CA.1992, pp.179-188.
- [10] P.Bosc.D.Dubois,and H.Prade, "Fuzzy functional dependencies and redundancy elimination." J.Amer. Soc.inform. Sci.,vol.49.pp.217-235,1998