Fuzzy Querying Based on Relational Database

Anupriya[1], Prof. Rahul Rishi[2]

[1](Department of CSE, Bhagat Phool Singh Mahila Vishwavidyalaya , Sonipat, India)
[2](Department of Computer Science & Engg, UIET (Maharshi Dayanand University), Rohtak, India)

Abstract:-*The traditional query in relational database is unable to satisfy the needs for dealing with fuzzy linguistic values. In this paper, a new data query technique composed of fuzzy theory and MS-SQL is provided. Here, the query can be implemented for fuzzy linguistic variables query via an interface to Microsoft ASP.NET. It is being applied to an realistic instance i.e. questions could be expressed by fuzzy linguistic values such as young age, high blood pressure, average heart beat etc in Patients' relational database. This could be widely used to realize the other fuzzy query based on database.*
Keywords: *ASP.NET, Fuzzy query, Fuzzy theory, MS-SQL, Relational database.*

I.   Introduction

Database management systems (DBMS) are widely used software products in many kinds of systems. As we know natural language is the main communication method for humans, but this causes it difficult to handle imperfect information. Imperfect information [1][2] can be inconsistent, imprecise, ambiguous, uncertain or vague. However, the Complexity is limited in precise data processing and is unable to directly express fuzzy concepts of natural language. For instance, in patients' relational database, to deal with a query statement like "young age and good heartbeat or low blood sugar ", it is difficult to construct SQL because the query words are fuzzy expressions. In order to obtain query results, there are two basic methods of research in the use of SQL Combined fuzzy theory in DBMS. The first is still to build a classic relation database, only to modify or extend SQL query by transforming query conditions to a fuzzy scope. After that, change it to precise SQL clause. This procedure is easy and consistent with ordinary query, but lacks flexibility. Sometime it is apt to produce query errors. The second approach is to assume that the database is fuzzy sets and fuzzy logic is used to make it easier and more human consistent. This is mainly done by constructing a database model based on fuzzy logic. When designing this database and modify its data structure, many tables including fuzzy fields' values may be added.These tables could be transformed from row tables. For instance, in the patients' relational database "Age", "Blood pressure", and "heart beat" etc. are fuzzy fields. When data is being inputted, the precise data may be transformed to fuzzy data and stored in database.

This paper deals with the second method. At beginning, consider a conventional patients database. Then, consider the second method to cover the fuzzy query in patient's database. Third, present how to define fuzzy elements appearing in the query, mainly the transformation of a fuzzy query into its equivalent legitimate MS- SQL query.

II.   Fuzzy Expressions In Patient Information Database

The Patient Information Database is one of the most important and basic data sources. In ordinary circumstances, "Append, Delete and Update" are common and important data manipulations. As we all know, it is not realistic to extract partly available data records from huge database. So, if an effective query method is provided, it may be more convenient for us to quickly locate useful records to process.

In this Patient Information database, ID field may be used as Primary key. The other query fields could be any query index which can be combined to query conditions. Sometimes, the query information is unable to set relations directly. For example, young age and good heartbeat or low blood sugar" information. By SQL, precise fields cannot be located. To solve these types of queries, we present the method which combines fuzzy theory with SQL.

2.1 Fuzzy Sets

Fuzzy sets[4] are extension and generalization of the basic concepts of crisp sets. It allows partial membership. Suppose U is universe of discourse , F is the fuzzy set of terms , u is membership function . u [0,1] and F$\leq$ U .
In given Patient Information Database , Gender (Male , Female) is precise attribute while Age(Young , Middle , Old , Very Old) Bloodpressure (low , Good , high) , cholesterol(Good , High , Very high) , heartbeat rate (Excellent , good , Average , Poor) ,Smoking level (Low , middle , High) , Blood sugar(Normal , Average , High). So , there are 6 fuzzy attributes Input fields are . Output field retrieves the selected query entries.In this

study , we use systolic and diasystole blood pressure. About blood sugar, fasting value and Post meal are considered.

III.   Method

In this section, the fuzzy inference engine designing , membership functions[5] are included.

(i) *Fuzzy inference engine designing*: Most important application of fuzzy system is in uncertain issues. When any query is having imperfect information, fuzzy logic is a suitable tool for dealing with that query. So, first step of fuzzy inference engine designing is determination of input and output variables. After that membership function for all attributes will be designed.

(ii) *Input attributes* :various input attributes[6] are :

1. **Gender**: This input field is divided into two sets (Male, Female).
2. **Age** : This input field is divided into 4 fuzzy sets (Young , Middle , Old , Very Old)
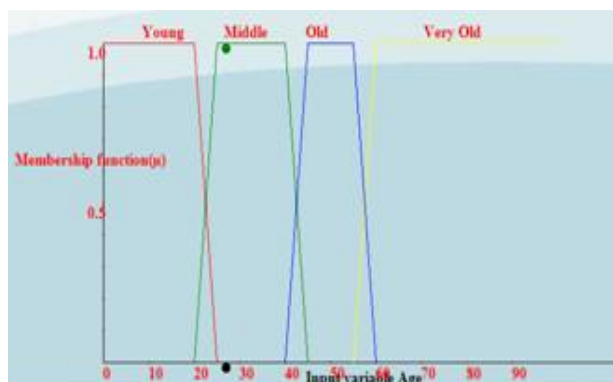
Input field	Fuzzy Set	Range
Age	Young	< 25
	Middle	20-45
	Old	40-60
	Very Old	>55

Table 1   FIG1. Membership functions of Age

Membership functions of these fuzzy sets are of "trapezoidal" type.

$$\mu_{young}(x) = \begin{cases} 1 & x \leq 20 \\ (25-x)/5 & 20 < x\, 25 \end{cases}$$

$$\mu_{middle}(x) = \begin{cases} 1 & 25 \leq x \leq 40 \\ (45-x)/5 & 40 < x < 45 \\ (x-20)/5 & 20 < x < 25 \end{cases}$$

$$\mu_{old}(x) = \begin{cases} 1 & 45 \leq x \leq 55 \\ (60-x)/5 & 55 < x < 60 \\ (x-40)/5 & 40 < x < 45 \end{cases}$$

$$\mu_{veryold}(x) = \begin{cases} 1 & x \geq 60 \\ (x-55)/5 & 55 < x < 60 \end{cases}$$

$$(1)$$

3. **Blood Pressure** : In this field, both systolic and diastolic blood pressure are considerd . This input variable has divided to 3 fuzzy sets. Fuzzy sets are "Low", "Good", "High" and their membership functions are trapezoidal. Fuzzy membership expressions for blood pressure input field are defined in (Eq.(2)). These fuzzy sets will be shown in Table 2. Membership functions of blood pressure field is shown in FIG.2.

Table 2

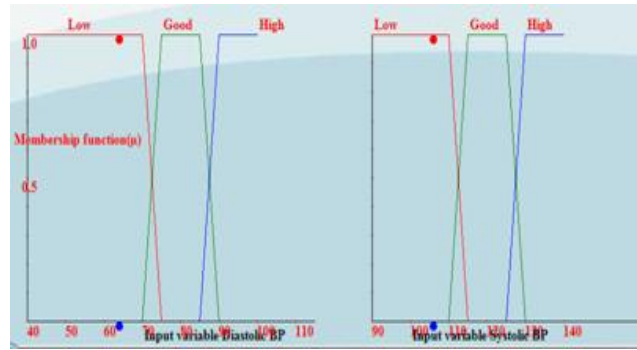| Input field	Fuzzy Set	Range	
		Diastolic	Systolic
Blood Pressure	Low	< 70	<115
	Good	70-90	115-130
	High	>85	>125

FIG 2. Membership functions of Blood Pressure

In case of Diastolic BP, a = 70*; b = 75; c = 85; d* = 90 while in case of Systolic BP, *a* = 110*; b* =115*; c* = 125*; d* = 130.

$$\mu_{low}(x) = \begin{cases} 1 & x \leq a \\ (b-x)/b-a & a < x < b \end{cases}$$

$$\mu_{Good}(x) = \begin{cases} 1 & b \leq x \leq c \\ (d-x)/d-c & c < x < d \\ (x-a)/b-a & a < x < b \end{cases}$$

$$\mu_{high}(x) = \begin{cases} 1 & x \geq d \\ (x-c)/d-c & c < x < d \end{cases} \tag{2}$$

Heart beat Rate: In this field, 4 linguistic variables (fuzzy sets) (Excellent, Good, Average and Poor) are considered. In Table 3, these fuzzy sets are defined. Membership functions are trapezoidal that will be shown in FIG. 3. Eq. (3) shows fuzzy membership function expressions.

Input field	Fuzzy Set	Range
	Excellent	49-65
Heart Beat Rate	Good	61-75
	Average	67-82
	Poor	>80

Table 3
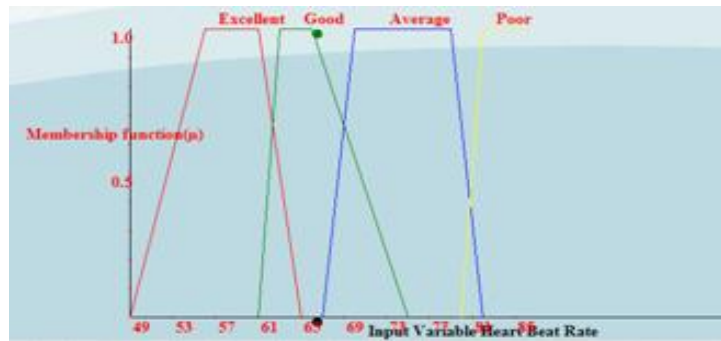
Fig 3 Membership functions of Heartbeat Rate

$$\mu_{excellent}(x) = \begin{cases} 1 & 56 \leq x \leq 61 \\ (65-x)/4 & 61 < x < 65 \\ (x-49)/7 & 49 < x < 56 \end{cases}$$

$$\mu_{Good}(x) = \begin{cases} 1 & 63 \leq x \leq 66 \\ (75-x)/9 & 66 < x < 75 \\ (x-61)/2 & 61 < x < 63 \end{cases}$$

$$\mu_{average}(x) = \begin{cases} 1 & 70 \leq x \leq 79 \\ (82-x)/3 & 79 < x < 82 \\ (x-67)/3 & 67 < x < 70 \end{cases}$$

$$\mu_{poor}(x) = \begin{cases} 1 & x \geq 82 \\ (x-80)/2 & 80 < x < 82 \end{cases} \qquad (3)$$

4. **Cholesterol** : Cholesterol field has 3 fuzzy sets (Good, High and Very high). These fuzzy sets have been shown in Table 4. Membership functions are Trapezoidal which will be shown in FIG.4. Eq. (4) shows these membership function expressions of cholesterol.

Cholesterol		
	Good	49-65
	High	61-75
	Very High	67-82

Table 4

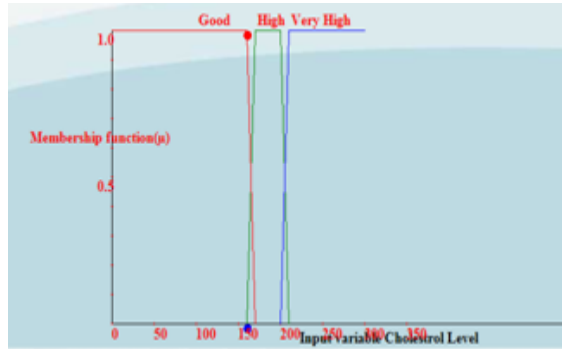FIG 4. Membership functions of Cholesterol

$$\mu_{good}(x) = \begin{cases} 1 & x \leq 160 \\ (160-x)/10 & 160 < x < 170 \end{cases}$$

$$\mu_{High}(x) = \begin{cases} 1 & 170 \leq x \leq 200 \\ (210-x)/10 & 200 < x < 210 \\ (x-160)/10 & 160 < x < 170 \end{cases}$$

$$\mu_{veryhigh}(x) = \begin{cases} 1 & x \geq 210 \\ (x-200)/10 & 200 < x < 210 \end{cases} \qquad (4)$$

5. Smoke level: Smoke field defines the number of cigarettes patient takes per day. It has 3 fuzzy sets (Low, Middle and High). These fuzzy sets have been shown in Table 5. Membership functions are Trapezoidal which will be shown in FIG.5. Eq. (5) shows these membership function expressions of smoke level.

Input field	Fuzzy Set	Range
Smoke Level	Low	< 4
	Middle	2-8
	High	>6

Table 5
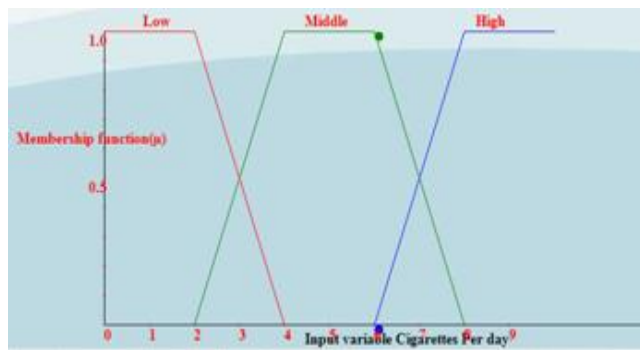

FIG.5 Membership functions of Smoke Level

$$\mu_{good}(x) = \begin{cases} 1 & x \leq 2 \\ (4-x)/2 & 2 < x < 4 \end{cases}$$

$$\mu_{middle}(x) = \begin{cases} 1 & 4 \leq x \leq 6 \\ (8-x)/2 & 6 < x < 8 \\ (x-2)/2 & 2 < x < 4 \end{cases}$$

$$\mu_{high}(x) = \begin{cases} 1 & x \geq 8 \\ (x-6)/2 & 6 < x < 8 \end{cases} \qquad (5)$$

6. Blood Sugar (Diabetes): In this field, Fasting value and Post meal blood sugar values are considered. This input variable has divided to 3 fuzzy sets. Fuzzy sets are "Normal", "Average", "High" and their membership functions are trapezoidal. Fuzzy membership expressions for blood sugar input field are defined in (Eq.(6)). These fuzzy sets will be shown in Table 6. Membership functions of blood pressure field will be shown in FIG.6.

Input field	Fuzzy Set	Range	
		FastingValue	PostMeal
Blood Sugar	Normal	70- 100	110-140
	Average	90-125	130-200
	High	>115	>190

Table 6

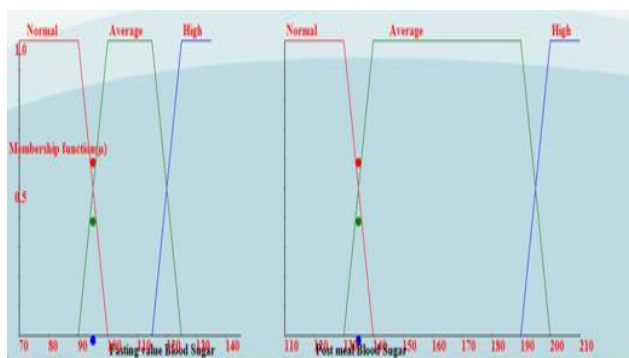FIG 6. Membership functions of Blood Sugar

In case of Fasting Value, *a* = 90*; b* = 100*; c* = 115*; d* = 125; e= 70 while in case of Posting meal, *a* = 130*; b* =140*; c* = 190*; d* = 200; e = 110.

$$\mu_{normal}(x) = \begin{cases} 1 & e \leq x \leq a \\ (b-x)/b-a & a < x < b \end{cases}$$

$$\mu_{average}(x) = \begin{cases} 1 & b \leq x \leq c \\ (d-x)/d-c & c < x < d \\ (x-a)/b-a & a < x < b \end{cases}$$

$$\mu_{high}(x) = \begin{cases} 1 & x \geq d \\ (x-c)/d-c & c < x < d \end{cases} \qquad (6)$$

7. Strength: Strength field asks about the different strength measures i.e. nominal, average or high .In order to search more valuable records in patient database, more words would be added for expressing the degree of information, for example: "nominal", "average", "high" etc.

(iii) Query Implementation

The method mentioned below is implemented mainly by Microsoft Visual Studio 8.0 combined SQL-SELECT query clause. The basic method used for query processing can be shown as in Fig.7. Initially, user will submit the query in natural language. Then in Fuzzy Interpreter phase, fuzzy attributes are identified from the given fuzzy query.
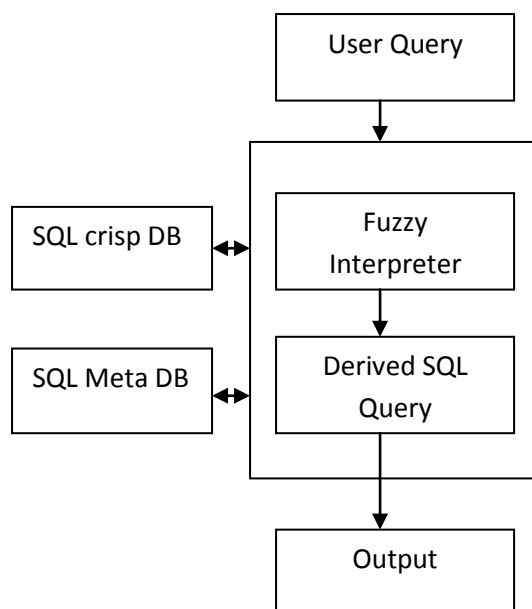
Fig. 7 Query Processing Architecture

Based on the fuzzy attribute fuzzy interpreter fetches the domain set of that attribute from the database and then finds the membership value for fuzzy equality of each domain value of the fuzzy attribute with respect to the given fuzzy data in the fuzzy query from fuzzy table. In fuzzy table, membership values for different fuzzy attributes are being calculated using the above mentioned formulas. If the query has more than one fuzzy attributes then in similar way it finds the membership value for fuzzy equality for other fuzzy attributes. In derived SQL query , based on the α-cut value[7], SQL syntax for the given fuzzy query will be generated. α-cut value is taken for considering the specific range of membership function values. Finally ,this query will be submitted for the desired results.

a) *Declaring attributes and defining fuzzy terms*
In this Patient database, one file is used to save the patients' personal information, such as :ID, Name, Gender, Age, Blood Pressure, heart beat , Diabetes , Cholesterol etc. In order to realize fuzzy query, the user should calculates the membership function values of the related attributes via programming codes at first. Then for the sake of causing the query conveniently, we save these values throughout putting them as a new table file , respectively. For example, the field AGE includes the most basic information(ID, Young, Middle, Old, Very Old) because of avoiding data redundancy. Except ID field, the other 4 fields are linked with each record's membership value. Also because of the relation between these table files, the values could input from this fuzzytable correctly and automatically. Similarly, other fuzzy attributes and their corresponding membership function values will also be defined as shown in Fig.8.

b) *Construct SQL query*
The SQL fuzzy query format may be expressed
like this:
    SELECT <attributes> FROM <tables> ;
    WHERE <fuzzy conditions>
Obviously, the fuzzy conditions clauses should be composed by common clauses and fuzzy ones. But in fact, it is could not be realized directly. Therefore, it must be transformed equivalence precise conditions linked with the fuzzy membership function values. The first step is to convert the user fuzzy query into a simplified classic SQL query as shown in Fig. 9.

SELECT * from Fuzzytable WHERE ((HbeatE >0.4 AND HbeatE <= 0.7) AND (CholH >0.7 AND CholH <=1.0))
The ideal SQL commands could be:
SELECT * FROM Fuzzytable WHERE HeartbeatRate ="average Excellent" AND Cholesterol= "high High";

	Table1: Query(anupriya...B.MDF)		Fuzzytable: Query(anup...B.MDF)		dbo.Fuzzytable: Table(...B.MDF)		combine.Designer.cs		try2.cs [Design]		try3.cs [Design]		
PatientId	AgeY	AgeM	AgeO	AgeVO	HbeatE	HbeatA	HbeatP	HbeatG	CholH	CholG	CholVH	BpL	BpH	BpA	BpL2
1	0	0.600...	0.4000...	0	0	0	0	1	0	1	0	1	0	0	1
2	0	0	0.2000...	0.8000...	1	0	0	0	1	0	0	0.20000...	0	0.800...	0
3	0	1	0	0	0.75	0	0	0.5	0.600...	0	0.4000...	0	0.300...	0.600...	0.600...
4	0.400...	0.600...	0	0	0.2857...	0	0	0	0.400...	0.6000...	0	0.40000...	0	0.600...	0
5	0	0.600...	0.4000...	0	0	0	0	1	0	1	0	1	0	0	1
6	0	1	0	0	0	0	0	1	0	1	0	1	0	0	1
7	1	0	0	0	0	0	0	1	0	1	0	1	0	0	0
8	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0
9	0	0	0.6000...	0.4000...	0	0.6666...	0	0	0	0	1	0.20000...	0	0.800...	0.200...
10	0	0	0	1	0	0.3333...	0	0.7777...	1	0	0	0.80000...	0	0.200...	0
11	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0
12	0	0	0	1	0	0.3333...	0.5	0	1	0	0	1	0	0	0.400...
13	0.200...	0.800...	0	0	0.8571...	0	0	0	0	1	0	1	0	0	0
14	0.400...	0.600...	0	1	0	0	0	1	0	0	0	1	0	0	0
15	0.600...	0.400...	0	1	0	0	0	1	0	0	0.60000...	0	0.400...	0	

Fig.8. Membership values file

Fig.9. Fuzzy query in Patient Information Database

IV. Conclusion

Our traditional database packages do not accept any fuzzy query statement. So, SQL for crisp query statement have to be provided. So, our aim in this paper is to deal with fuzzy queries. This paper mainly discusses the realization of fuzzy query through fuzzy theory and SQL combined C#. Also, a real life application of fuzzy query based on relational database (the Patient Information database) is provided. The definition of membership function used in this paper is flexible. It may be modified in different situations. As the development of intelligence information technology and intelligence database, fuzzy query and fuzzy data processing are surely be used in more widespread domain.

References

[1] Z.M. Ma,"A conceptual design methodology for fuzzy relational databases", Journal of Database Management , Vol. 16, No. 2 ,pp 66-83 , 2005.
[2] Anupriya and Rahul Rishi, " Review of Fuzzy Logical Database Models" , IOSR Journal of Computer Engineering , Vol. 8, No. 4, pp 24-30, 2013.
[3] E. F. Codd, "The relational model for database management", Version 2. Reading, MA: Addison-Wesley, 1990.
[4] L. A. Zadeh, " Fuzzy sets", Information and Control, Vol. 8, pp 338-353, 1965.
[5] J. Galindo, A. Urrutia and M. Piattini ," Fuzzy Databases: Modeling Design and Implementation" , IDEA Group Publishing, Hershey, USA , 2006.
[6] A. Adeli and M. Neshat, "A Fuzzy expert system for heart disease diagnosis", Proceedings of International MultiConference of Engineers and Computer Scientists(IMECS) ,Vol. 1, 2010
[7] P. Bosc and 0. Pivert, "On the efficiency of the alpha-cut distribution method to evaluate simple fuzzy relational queries", Advances in Fuzzy Systems Applications and Theory, Vol. 4, pp 251-260, 1995