INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
ISSN 1841-9836, 9(6):730-740, December, 2014.

Uncertain Query Processing using Vague Set or Fuzzy Set: Which One Is Better?

J. Mishra, S. Ghosh

Jaydev Mishra*
Computer Science and Engineering Department
College of Engineering and Management, Kolaghat
West Bengal-721171, India
*Corresponding author:jsm03@cemk.ac.in

Sharmistha Ghosh
Galgotias University, Greater Noida
Uttar Pradesh-201306, India
sharmisthag@yahoo.com

Abstract: In this paper we attempt to make a theoretical comparison between fuzzy sets and vague sets in processing uncertain queries. We have designed an architecture to process uncertain i.e. fuzzy or vague queries. In the architecture we have presented an algorithm to find the membership value that generates the fuzzy or vague representation of the attributes with respect to the given uncertain query. Next, a similarity measure is used to get each tuples similarity value with the uncertain query for both fuzzy and vague sets. Finally, a decision maker will supply a threshold or α-cut value based on which a corresponding SQL statement is generated for the given uncertain query. This SQL retrieves different result sets from the database for fuzzy or vague data. It has been shown with examples that vague sets give more accurate result in comparison with fuzzy sets for any uncertain query.
Keywords: uncertain data, similarity measures, fuzzy/vague interpreter.

1 Introduction

In the real world, vaguely specified data values appear in many applications such as sensor information, expert systems, decision analysis, medical sciences, management and engineering problems and so on. Fuzzy set theory has been proposed to handle such vagueness by generalizing the notion of membership in a set. Essentially, in a fuzzy set each element is associated with a point-value selected from the unit interval [0, 1], which is termed as the grade of membership in the set. A vague set, which is conceived as a further generalization of fuzzy set, uses the idea of interval-based membership instead of point-based membership as in the case of fuzzy sets. The interval-based membership in vague sets is more expressive in capturing vagueness of data.

Relational database systems have been extensively studied worldwide since Codd [1] had proposed the relational data model in 1970. Based on this model, several commercial relational database systems are available (see [2]- [4]). This data model usually takes care of precisely defined and unambiguous data. However, in the real world applications data are often partially known i.e., incomplete or imprecise. For example, instead of specifying that the height of David is 188 cm, one may say that the height of David is around 190 cm, or simply that David is tall. Other examples on uncertain data may be "Salaries of almost equally experienced employees are more or less the same" etc. All these are informative statements that may be useful in answering queries or making inferences. However, such type of data cannot be represented in the classical relational data model. In order to incorporate imprecise or uncertain data, the classical relational data model has been extended by several authors on the mathematical framework of fuzzy set theory which was initially introduced by Zadeh [5] in 1965. Based on this fuzzy set theory, various fuzzy relational database models, such as similarity-based relational model [6], possibility-based relational model [7] and some types of hybrid data models [8] have been proposed to model fuzzy information in relational databases. However, the most important issue in the utilization of any database system lies in its ability to process information and queries correctly. Several authors [9]- [12] have contributed to provide a theoretical contribution to query language for a fuzzy database model. In particular, Bosc et al. [11] and Nakajima [12] have extended the well known SQL language in the framework of fuzzy set theory and have developed a fuzzy SQL language, called SQLf.

It is believed that vague sets, proposed by Gau et al. [13] in 1993, that use interval-based membership values have more powerful ability to process imprecise information than traditional fuzzy sets. Thus the notion of vague sets has also been incorporated into relations in [14] and a vague SQL (VSQL) has been described. The VSQL allows the users to formulate a wide range of queries that occur in different modes of interaction between vague data and queries. In [15], Zhao and Ma have proposed a vague relational database model which is an extension of the classical relational model. Based on the proposed model and the semantic measure of vague sets, they have also investigated vague querying strategies and have given the form of vague querying with SQL.

In this paper, we have made an attempt to make a theoretical comparison between fuzzy sets and vague sets in processing uncertain queries. Firstly, we have designed an architecture to test uncertain queries. Next, we have presented an algorithm to retrieve membership values for imprecise data represented by fuzzy or vague sets. A similarity measure is then used to calculate each tuple's similarity value with the uncertain query for both fuzzy and vague sets. Finally, the decision maker provides a threshold value or α-cut based on which a corresponding SQL statement is generated for the given uncertain query. This SQL retrieves different result sets from the database for fuzzy data or vague data. In the present study, we have considered an Employee database and processed some uncertain queries using fuzzy data as well as vague data. Each time it has been observed that vague sets give more accurate result in comparison to fuzzy sets.

The rest of the paper is organized as follows. Section 2 presents some basic definitions related to fuzzy and vague sets. Similarity measure between two vague data is also defined in the same section. Section 3 represents an architecture for processing uncertain queries. In Section 4, an algorithm has been designed to get the appropriate membership value and represent domain value of fuzzy or vague attributes into fuzzy form or vague form. Section 5 establishes that a vague set is more appropriate than fuzzy set with real life examples. The concluding remarks appear in Section 6.

2 Basic Definitions

In this section, we introduce some basic concepts related to fuzzy and vague sets and similarity measure of two vague sets which have been utilized throughout the paper. Let U be the universe of discourse where an element of U is denoted by u.

2.1 Fuzzy Set

Definition 1. A Fuzzy set F in the universe of discourse U is characterized by a membership function $\mu_F : U \to [0, 1]$ and is defined as a set of ordered pairs $F = \{\langle u, \mu_F(u)\rangle : u \in U\}$ where $\mu_F(u)$ for each $u \in U$ denotes the grade of membership of u in the fuzzy set F.

2.2 Vague Set

Definition 2. A vague set V in the universe of discourse U is characterized by two membership functions given by:
slowromancapi@. a truth membership function $t_V : U \to [0, 1]$ and
slowromancapii@. a false membership function $f_V : U \to [0, 1]$,
where $t_V(u)$ is a lower bound of the grade of membership of u derived from the 'evidence for u', and $f_V(u)$ is a lower bound on the negation of u derived from the 'evidence against u', and $t_V(u) + f_V(u) \leq 1$. Thus the grade of membership $\mu_V(u)$ of u in the vague set V is bounded by a subinterval $[t_V(u), 1 - f_V(u)]$ of $[0, 1]$, i.e., $t_V(u) \leq \mu_V(u) \leq 1 - f_V(u)$. Then, the vague set V is written as $V = \{\langle u, [t_V(u), 1 - f_V(u)]\rangle : u \in U\}$. Here, the interval $[t_V(u), 1 - f_V(u)]$ is said to be the vague value to the object u and is denoted by $V_V(u)$.

For example, in disease diagnosis process of a medical system, the vague value $[0.3, 0.6]$ can be interpreted as "the report of disease in favour is 30%, against is 40% and another 30% is indeterminable". The precision of knowledge about u is clearly characterized by the difference $(1-f_V(u)-t_V(u))$. If this is small, then the knowledge about u is relatively precise. However, if it is large, we know correspondingly little. If $t_V(u)$ is equal to $(1 - f_V(u))$, the knowledge about u is precise, and vague set theory reverts back to fuzzy set theory. If $t_V(u)$ and $(1 - f_V(u))$ are both equal to 1 or 0, depending on whether u does or does not belong to V, the knowledge about u is exact and the theory goes back to that of ordinary set. Thus any crisp or fuzzy set may be considered as a special case of vague sets.

2.3 Similarity Measure

There have been some studies in literature which discuss the topic concerning how to measure the degree of similarity between vague sets [16]- [19]. In [19] it was pointed out by Lu et al. that the similarity measures defined in [16]- [18] did not fit well in some cases. They have proposed a new similarity measure between vague sets which turned out to be more reasonable in more general cases. The same has been used in the present work which is defined as follows:

Definition 3. Similarity Measure between two vague values
Let x and y be any two vague values such that $x = [t_x, 1 - f_x]$ and $y = [t_y, 1 - f_y]$, where $0 \leq t_x \leq 1 - f_x \leq 1$, and $0 \leq t_y \leq 1 - f_y \leq 1$. Let $SE(x, y)$ denote the similarity measure between x and y. Then

$$SE(x, y) = \sqrt{(1 - (|(t_x - t_y) - (f_x - f_y)|/2))\,(1 - |(t_x - t_y) + (f_x - f_y)|)}.$$

Definition 4. Similarity Measure between two vague sets
Let $U = \{u_1, u_2, u_3, \cdots, u_n\}$ be the universe of discourse. Let A and B be two vague sets on U, such that $A = \{< u_i, [t_A(u_i), 1 - f_A(u_i)] >, \forall u_i \in U\}$, where $t_A(u_i) \leq \mu_A(u_i) \leq 1 - f_A(u_i)$ and $1 \leq i \leq n$. $B = \{< u_i, [t_B(u_i), 1 - f_B(u_i)] >, \forall u_i \in U\}$, where $t_B(u_i) \leq \mu_B(u_i) \leq 1 - f_B(u_i)$ and $1 \leq i \leq n$. Now, the similarity measure between A and B, denoted by $SE(A, B)$ is defined as:

$$SE(A, B) = \frac{1}{n}\sum_{i=1}^{n} SE\left([t_A(u_i), 1 - f_A(u_i)], [t_B(u_i), 1 - f_B(u_i)]\right) = \frac{1}{n}\sum_{i=1}^{n}\sqrt{P * Q}$$

where $P = (1 - (|(t_A(u_i) - t_B(u_i)) - (f_A(u_i) - f_B(u_i))|/2))$

and $Q = (1 - |(t_A(u_i) - t_B(u_i)) - (f_A(u_i) - f_B(u_i))|)$

3 Architecture for Processing Uncertain Query

Below is the architecture for processing an imprecise or uncertain query.

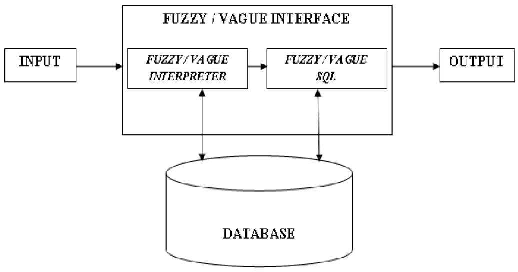

Figure 1: Uncertain Query Processing Architecture

Working principle of all components of the above proposed architecture is given below:

Input: A relational database with fuzzy or vague attributes, uncertain query, threshold value or α-tolerance value given by Decision Maker.
Fuzzy/Vague Interface:
It has two components, namely, **Fuzzy/Vague Interpreter** and **Fuzzy/Vague SQL**.
Fuzzy Interpreter: In this phase, the fuzzy attributes as well as the fuzzy data are identified from the given fuzzy query. Next, the fuzzy interpreter represents all domain values of each of the fuzzy attributes. **Algorithm 1** presented below in section 4 is then used to determine membership value for each domain value of all the fuzzy attributes. This gives us the fuzzy representation of the attributes with respect to the fuzzy data identified from the given query. The above fuzzy representation is then converted to a corresponding vague form whose truth membership value is same as the membership calculated in the fuzzy representation and false membership value is (1-truth membership value). After that a suitable similarity measure formula is to be used to measure the similarity between vague representation of each fuzzy attribute and the corresponding vague representation of the relevant fuzzy data given in the uncertain query. The same method will be applied for all fuzzy attributes appearing in the fuzzy query. If the query has more than one fuzzy attribute, then the similarity measure of tuples is obtained as the intersection of the similarity measures for each attribute.
Vague Interpreter: In this case, the vague attributes and the vague data are identified from the given uncertain query. Next, the vague interpreter will represent all domain values of each of the vague attributes. **Algorithm 1** is now used to get the truth membership values of the vague attributes while the decision maker will supply the false membership values with the condition that the sum of truth and false membership values should not exceed 1. After that, as before, a similarity measure formula is used to measure the similarity between each vague representation of a domain value and vague data given in the query. If the query has more than one vague attribute, once again the intersection of the similarity measures for individual attributes will give the similarity measure of tuples.
Fuzzy/Vague SQL: In this phase, the decision maker will supply a threshold value or α-cut value based on which a corresponding SQL statement for the given uncertain i.e., fuzzy or vague query will be generated.
Output: Finally, the SQL generated above is submitted to the database to get the desired result.

4 Algorithm for finding Membership Value

The following algorithm finds the membership value for each domain value of fuzzy or vague attributes with respect to fuzzy or vague data given in the uncertain query.

Algorithm 1 Membership value calculation

Input: Fuzzy/Vague attributes and Fuzzy/Vague data given in the uncertain query.
Output: A membership value in the interval [0,1].
Method: First find the fuzzy/vague attributes from the fuzzy/vague query.
for each fuzzy/vague attribute **do**
begin
fdata ← data value for the fuzzy/vague attribute of the query
range = maxDomainValue - minDomainValue
avg ← mean value of the domain set of the fuzzy/vague attribute
B ← avg
while(avg ≤ range) **do**
begin
avg = avg + B
end while loop
for each tuple of the relation **do**
begin
tupleValue ← corresponding tuple value from the fuzzy/vague attribute domain
membershipValue = 1-(|fdata-tupleValue| / avg)
end for loop of tuple
end for loop of fuzzy/vague attribute

5 Vague Sets have an extra edge over Fuzzy Sets

In this section, we have experimentally shown with real life examples that vague sets give more accurate result than fuzzy sets. To illustrate this fact, we have considered the following Employee EMP relational database:

Table 1: EMP Relation

Name	Age (yrs)	Exp (yrs)	Sal (Rs)
Prof. Smith	25	1	20000
Prof. Ganguly	52	25	55000
Prof. Roy	38	15	38000
Prof. David	48	23	53000
Prof. Maity	34	10	32000
Prof. Das	30	4	27000
Prof. Ahuja	50	26	55500
Prof. Sharma	51	16	40000
Prof. Kundu	45	22	50000
Prof. Dutta	54	33	80000

Next we consider following uncertain queries to explain that vague sets give better result in comparison to fuzzy sets.

Uncertain query 1: "Find the details of the Professors whose age is around 50".

i) **Solution with Fuzzy Sets:** In the above *uncertain query 1*, fuzzy attribute is ***Age*** and fuzzy data is ***around 50***. Now, we apply our **algorithm 1** to get the membership value corresponding to each domain value of fuzzy attribute ***Age***.
Input: Algorithm needs the following two inputs: fuzzy attribute ***Age*** and fuzzy data ***around 50***.
Method: Calculation of membership value for each tuple value of fuzzy attribute ***Age*** based on fuzzy data ***around 50***.
dom(Age) = $\{25, 52, 38, 48, 34, 30, 50, 51, 45, 54\}$
given fdata=50
range = 54 - 25 = 29
Avg = 42.7
B = 42.7
Avg$\geq$ range then Avg remain same i.e., Avg = 42.7
Now, we need to find the membership value using the formula specified in the **algorithm 1**:
$membershipValue = 1 - (|fdata - tupleValue|/Avg)$
for the 1st tuple : $membershipValue = 1 - (|50 - 25|/42.7) = 0.41$
for the 2nd tuple : $membershipValue = 1 - (|50 - 52|/42.7) = 0.95$
for the 3rd tuple: $membershipValue = 1 - (|50 - 38|/42.7) = 0.72$
for the 4th tuple: $membershipValue = 1 - (|50 - 48|/42.7) = 0.95$
for the 5th tuple: $membershipValue = 1 - (|50 - 34|/42.7) = 0.63$
for the 6th tuple: $membershipValue = 1 - (|50 - 30|/42.7) = 0.53$
for the 7th tuple: $membershipValue = 1 - (|50 - 50|/42.7) = 1$
for the 8th tuple: $membershipValue = 1 - (|50 - 51|/42.7) = 0.98$
for the 9th tuple: $membershipValue = 1 - (|50 - 45|/42.7) = 0.88$
for the 10th tuple: $membershipValue = 1 - (|50 - 54|/42.7) = 0.91$

The fuzzy representation of the EMP relation w.r.t. **uncertain query 1** is now depicted in Table 2. In particular, the fuzzy representation of the attribute ***Age*** appears in third column of the Table 2. Next in fourth column, we have shown the corresponding vague representation of these fuzzy data. These vague values are then used to find the similarity measures (**S.M.**) with fuzzy data **around 50** whose vague representation is $< 50, [1, 1] >$. The similarity measures have been calculated using the same formula as presented in **definition 3**. For example, consider the following two vague data:
$x =< 50, [1, 1] >$ and $y =< 25, [0.41, 0.41] >$. Here $t_x = 1$, $f_x = 0$, $t_y = 0.41$, $f_y = 0.59$.
Then, $SE(x, y) = \sqrt{(1 - (|(1 - 0.41) - (0 - 0.59)|/2))\,(1 - |(1 - 0.41) + (0 - 0.59)|)}$
$= \sqrt{(1 - 0.59)} = \sqrt{0.41} = 0.64$

Again, for the vague values $x =< 50, [1, 1] >$ and $y =< 52, [0.95, 0.95] >$, $t_x = 1$, $f_x = 0$, $t_y = 0.95$, $f_y = 0.05$.
Then, $SE(x, y) = \sqrt{(1 - (|(1 - 0.95) - (0 - 0.05)|/2))\,(1 - |(1 - 0.95) + (0 - 0.05)|)}$
$= \sqrt{(1 - 0.05)} = \sqrt{0.95} = 0.98$ and so on.
Using the notation: FD = Fuzzy Data, VD = Vague Data, we represent in Table 2:

Now, if the threshold value or α-cut given by the decision maker is 0.95, then the corresponding SQL statement of the uncertain query 1 is generated as below:
Select $*$ *from EMP where S.M.(tuple)*$\geq$*0.95* which retrieves the following resultant tuples given

Table 2: Fuzzy Representation of EMP Relation w.r.t Uncertain Query 1

Name	Age	Fuzzy **Age** with FD **around 50**	Vague FD **Age**	S.M. with VD $< 50, [1, 1] >$	Exp	Sal	S.M. (tuple)
Prof. Smith	25	$< 25, .41 >$	$< 25, [.41, .41] >$	.64	1	20000	.64
Prof. Ganguly	52	$< 52, .95 >$	$< 52, [.95, .95] >$	.98	25	55000	.98
Prof. Roy	38	$< 38, .72 >$	$< 38, [.72, .72] >$	.85	15	38000	.85
Prof. David	48	$< 48, .95 >$	$< 48, [.95, .95] >$	.98	23	53000	.98
Prof. Maity	34	$< 34, .63 >$	$< 34, [.63, .63] >$	.79	10	32000	.79
Prof. Das	30	$< 30, .53 >$	$< 30, [.53, .53] >$	.73	4	27000	.73
Prof. Ahuja	50	$< 50, 1 >$	$< 50, [1, 1] >$	1	26	55500	1
Prof. Sharma	51	$< 51, .98 >$	$< 51, [.98, .98] >$	.99	16	40000	.99
Prof. Kundu	45	$< 45, .88 >$	$< 45, [.88, .88] >$	.94	22	50000	.94
Prof. Dutta	54	$< 54, .91 >$	$< 54, [.91, .91] >$	.95	33	80000	.95

in Table 3 from the EMP database in Table 2.

Table 3: Resultant Relation of Uncertain Query 1 for Fuzzy Set at Threshold value α=0.95

Name	***Age***	***Exp***	***Sal***
Prof. Ganguly	52	25	55000
Prof. David	48	23	53000
Prof. Ahuja	50	26	55500
Prof. Sharma	51	16	40000
Prof. Dutta	54	33	80000

ii) **Solution with Vague Sets:** Next, we process the same *uncertain query 1* for vague sets. Here, vague attribute is ***Age*** and vague data is ***around 50***.

It is then necessary to represent all domain values of attribute ***Age*** into vague form whose truth membership values are calculated from the **algorithm 1** and false membership values are provided by the decision maker considering the restriction that sum of truth and false membership values ≤1. Similarity measures are then calculated using the same formula as used for fuzzy attributes.

Let us consider the two vague data $x =< 50, [1, 1] >$ and $y =< 25, [0.41, 0.5] >$. Here $t_x = 1$, $f_x = 0$, $t_y = 0.41$, $f_y = 0.5$.

Then, $SE(x, y) = \sqrt{(1 - (|(1 - 0.41) - (0 - 0.5)|/2)) \, (1 - |(1 - 0.41) + (0 - 0.5)|)}$
$= \sqrt{(1 - 1.09/2) * (1 - 0.09)} = \sqrt{0.455 * 0.91} = \sqrt{0.41405} = 0.64$

Again, for $x =< 50, [1, 1] >$ and $y =< 52, [0.95, 0.98] >$, $t_x = 1$, $f_x = 0$, $t_y = 0.95$, $f_y = 0.02$.
Then, $SE(x, y) = \sqrt{(1 - (|(1 - 0.95) - (0 - 0.02)|/2)) \, (1 - |(1 - 0.95) + (0 - 0.02)|)}$
$= \sqrt{(1 - 0.07/2)(1 - 0.03)} = \sqrt{0.93605} = 0.97$ and so on.

The Table 4 given below shows the complete vague representation of EMP relation w.r.t. the uncertain query 1.

with the same threshold or α-cut value 0.95, the following SQL statement of the uncertain query 1 for vague set will be generated:

Select * from EMP where S.M.(tuple)≥0.95

which now retrieves the following resultant tuples given in Table 5 from the EMP database for

Table 4: Vague Representation of EMP Relation w.r.t Uncertain Query 1

Name	Age	Vague representation of vague data ***Age***	S.M. with vague data $< 50, [1, 1] >$	Exp	Sal	S.M. (tuple)
Prof. Smith	25	$< 25, [.41, .5] >$	.64	1	20000	.64
Prof. Ganguly	52	$< 52, [.95, .98] >$	.97	25	55000	.97
Prof. Roy	38	$< 38, [.72, .8] >$	.84	15	38000	.84
Prof. David	48	$< 48, [.95, .98] >$	.97	23	53000	.97
Prof. Maity	34	$< 34, [.63, .75] >$	.78	10	32000	.78
Prof. Das	30	$< 30, [.53, .7] >$	.71	4	27000	.71
Prof. Ahuja	50	$< 50, [1, 1] >$	1	26	55500	1
Prof. Sharma	51	$< 51, [.98, 1] >$	.98	16	40000	.98
Prof. Kundu	45	$< 45, [.88, .93] >$	.93	22	50000	.93
Prof. Dutta	54	$< 54, [.91, .95] >$	.94	33	80000	.94

vague set in Table 4.

Table 5: Resultant Relation of Uncertain Query 1 for Vague Set at threshold value α=0.95

Name	***Age***	***Exp***	***Sal***
Prof. Ganguly	52	25	55000
Prof. David	48	23	53000
Prof. Ahuja	50	26	55500
Prof. Sharma	51	16	40000

It may be noted from Tables 3 and 5 that vague set gives better solution than fuzzy set since the SQL statement with vague query does not retrieve the tuple of Prof. Dutta with age 54 that has been fetched with the fuzzy query. It may be observed that 54 is less closer to 50 compared to the values of the attribute ***Age*** in all the other tuples retrieved by the SQL statement.

Next, we consider an uncertain query where more than one attribute are fuzzy or vague in nature. **Uncertain query 2**: "Find the details of the Professors whose age is **around** 50 and experience is **more or less** 25".

i) Solution with Fuzzy Sets: **Uncertain query 2** has two fuzzy attributes, ***Age*** and ***Experience***. Applying **algorithm 1** and **definition 3** for calculating membership values and similarity measures respectively, we get the following fuzzy representation of EMP relation (FD = Fuzzy Data):

Here, μ_1 denotes the similarity measures of ***Age*** attribute with respect to FD ***Age*** around 50 [for detail calculation **see** Table 2],
μ_2 denotes the similarity measures with respect to fuzzy attribute ***Exp***,
and $\mu = \mu_1 \cap \mu_2$ denotes the similarity measures of tuples.

Then, the result is tested for different **threshold** or α-cut values given by the decision maker.

Table 6: Fuzzy Representation of EMP Relation w.r.t Uncertain Query 2

Name	Age	μ_1	Exp	Fuzzy **Exp** with FD **almost 25**	Vague FD **Exp**	μ_2	Sal	μ
Prof. Smith	25	.64	1	$< 1, .3 >$	$< 1, [.3, .3] >$	.56	20000	.56
Prof. Ganguly	52	.98	25	$< 25, 1 >$	$< 25, [1, 1] >$	1	55000	.98
Prof. Roy	38	.85	15	$< 15, .71 >$	$< 15, [.71, .71] >$	.84	38000	.84
Prof. David	48	.98	23	$< 23, .94 >$	$< 23, [.94, .94] >$	.97	53000	.97
Prof. Maity	34	.79	10	$< 10, .5 >$	$< 10, [.5, .5] >$	.71	32000	.71
Prof. Das	30	.73	4	$< 4, .4 >$	$< 4, [.4, .4] >$	.63	27000	.63
Prof. Ahuja	50	1	26	$< 26, .97 >$	$< 26, [.97, .97] >$	.99	55500	.99
Prof. Sharma	51	.99	16	$< 16, .74 >$	$< 16, [.74, .74] >$	.86	40000	.86
Prof. Kundu	45	.94	22	$< 22, .91 >$	$< 22, [.91, .91] >$	.95	50000	.94
Prof. Dutta	54	.95	33	$< 33, .77 >$	$< 33, [.77, .77] >$	.88	80000	.88

Case a) for α=0.95, the SQL statement is *Select* $*$ *from EMP where* $\alpha \geq$*0.95* which retrieves the resultant Table 7 from Table 6 as follows:

Table 7: Resultant Relation of Uncertain Query 2 for Fuzzy Set at threshold value α=0.95

Name	**Age**	**Exp**	**Sal**
Prof. Ganguly	52	25	55000
Prof. David	48	23	53000
Prof. Ahuja	50	26	55500

Case b) for α=0.87, the SQL statement is *Select* $*$ *from EMP where* $\alpha \geq$*0.87* and the resultant table is shown below in Table 8:

Table 8: Resultant Relation of Uncertain Query 2 for Fuzzy Set at threshold value α=0.87

Name	**Age**	**Exp**	**Sal**
Prof. Ganguly	52	25	55000
Prof. David	48	23	53000
Prof. Ahuja	50	26	55500
Prof. Kundu	45	22	50000
Prof. Dutta	54	33	80000

ii) Solution with Vague Sets:
Again, using **algorithm 1** and **definition 3**, the vague representation of EMP relation for uncertain query 2 may be obtained as follows (VD= Vague Data):

The result is now tested for vague set with the same threshold or α-cut values.
Case a) for α=0.95, the SQL statement is *Select* $*$ *from EMP where* $\mu \geq$*0.95* which retrieves from Table 9 the following resultant table as
Case b) for α=0.87, SQL statement is *Select* $*$ *from EMP where* $\mu \geq$*0.87* and the resultant table is

From Tables 7 and 10 it may be observed that the resultant sets of the uncertain query 2 for both fuzzy data and vague data are same for the threshold value α=0.95.
However, when the same query is tested with α-cut value 0.87, Tables 8 and 11 show that the vague sets certainly gives better result than fuzzy sets because vague SQL has not retrieved the tuple Prof. Dutta with age 54 and experience 33 which is not so closer to age 50 and experience 25.

Table 9: Vague Representation of EMP Relation w.r.t Uncertain Query 2

Name	Age	Vague **Age** with VD **around 50**	μ_1	Exp	Vague **Exp** with VD **almost 25**	μ_2	Sal	μ
Prof. Smith	25	$< 25, [.41, .5] >$	.64	1	$< 1, [.3, .4] >$	.56	20000	.56
Prof. Ganguly	52	$< 52, [.95, .98] >$	.97	25	$< 25, [1, 1] >$	1	55000	.97
Prof. Roy	38	$< 38, [.72, .8] >$	.84	15	$< 15, [.71, .8] >$	.83	38000	.83
Prof. David	48	$< 48, [.95, .98] >$	.97	23	$< 23, [.94, .98] >$	.96	53000	.96
Prof. Maity	34	$< 34, [.63, .75] >$	.78	10	$< 10, [.5, .6] >$	.7	32000	.7
Prof. Das	30	$< 30, [.53, .7] >$	.71	4	$< 4, [.4, .43] >$	.63	27000	.63
Prof. Ahuja	50	$< 50, [1, 1] >$	1	26	$< 26, [.97, 1] >$	.98	55500	.98
Prof. Sharma	51	$< 51, [.98, 1] >$	.98	16	$< 16, [.74, .82] >$	.85	40000	.85
Prof. Kundu	45	$< 45, [.88, .93] >$	.93	22	$< 22, [.91, .96] >$	.94	50000	.94
Prof. Dutta	54	$< 54, [.91, .95] >$	.94	33	$< 33, [.77, .85] >$	.86	80000	.88

Table 10: Resultant Relation of Uncertain Query 2 for Vague Set at threshold value α=0.95

Name	***Age***	***Exp***	***Sal***
Prof. Ganguly	52	25	55000
Prof. David	48	23	53000
Prof. Ahuja	50	26	55500

Table 11: Resultant Relation of Uncertain Query 2 for Vague Set at threshold value α=0.87

Name	***Age***	***Exp***	***Sal***
Prof. Ganguly	52	25	55000
Prof. David	48	23	53000
Prof. Ahuja	50	26	55500
Prof. Kundu	45	22	50000

6 Conclusions

In this paper, we have proposed an architecture to process uncertain queries represented by fuzzy or vague data. We have also presented an algorithm that generates the fuzzy or vague representation of the attributes with respect to the given uncertain query. Similarity measure presented in definition 3 is used to find similarity measure of tuples w.r.t the given uncertain query in fuzzy or vague representation. Then the proposed architecture has been verified for uncertain queries using a real life example, both for the fuzzy as well as vague representation. In each case it has been observed that vague sets have produced more accurate result in comparison to fuzzy sets. Hence a vague relational database model may be more fruitful in processing real life data and queries than the conventional fuzzy data models. A DBMS that implements this vague set theoretic concept can thus become a more powerful software product than those currently available.

Bibliography

[1] Codd E. F. (1970); A Relational Model for Large Shared Data Banks, *Comm. of ACM*, 13(6): 377-387.

[2] Codd E. F. (1990); *The Relational Model for Database Management*, Addison Wesley.

[3] Date C. J. (2004); *An Introduction to Data Base Systems*, 8th ed., Addison Wesley.

[4] Elmasri R.; Navathe S. B. (2010); *Fundamentals of Database Systems*, 6th ed., Pearson.

[5] Zadeh L. A. (1965); Fuzzy Sets, *Information and Control*, 8(3): 338-353.

[6] Buckles P. B.; Petry F. E. (1982); A Fuzzy Representation of Data For Relational Databases, *Fuzzy Sets and Systems*, 7(3): 213-226.

[7] Raju K.V.S.V.N.; Majumdar A.K. (1988); Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database system, *ACM Transactions on Database Systems*, 13(2): 129-166.

[8] Ma Z. M.; Mili F. (2002); Handling fuzzy information in extended possibility-based fuzzy relational databases, *International Journal of Intelligent Systems*, 17(10): 925-942.

[9] Intan R.; Mukaidono M. (2000); Fuzzy functional dependency and its application to approximate data querying, *Proc. of international Database Engineering and Applications Symposium*, 47-54.

[10] Takahashi Y. (1993); Fuzzy database query languages and their relational completeness theorem, *IEEE Transactions on Knowledge and Data Engineering*, 5: 122-125.

[11] Bosc P.; Pivert O. (1995); SQLF: A relational database language for fuzzy querying, *IEEE Transaction on Fuzzy Systems*, 3(1): 1-17.

[12] Nakajima H. et al. (1993); Fuzzy Database Language and Library- Fuzzy Extension to SQL, *Second IEEE International Conference on Fuzzy Systems*, 1: 477-482.

[13] Gau W. L.; D. J. Buehrer (1993); Vague Sets, *IEEE Trans. Syst. Man, Cybernetics*, 23(2): 610-614.

[14] Lu A.; Ng W. (2005); Vague Sets or Intuitionistic Fuzzy Sets for Handling Vague Data: Which one is better?, *Lecture Notes in Computer Science*, 3716: 401-416.

[15] Zhao F.; Ma Z. M. (2009); Vague Query Based on Vague Relational Model, *AISC, Springer-Verlag Berlin Heidelberg*, 61: 229-238.

[16] Chen S. M. (1997); Similarity Measure between Vague Sets and between Elements, *IEEE Trans. Systems. Man and Cybernetics*, 27(1): 153-158.

[17] Hong D. H.; Kim C. (1999); A Note on Similarity Measures between Vague Sets and between Elements, *Information Sciences*, 115: 83-96.

[18] Li F.; Xu Z. (2001); Measures of Similarity between Vague Sets, *Journal of Software*, 12(6): 922-927.

[19] Lu A.; Ng W. (2004); Managing Merged Data by Vague Functional Dependencies, *LNCS, Springer-Verlag Berlin Heidelberg*, 3288: 259-272.