A Study of Vague Search to Answer Imprecise Query

Ashit Kumar Dutta [1]* and Ranjit Biswas [2]

[1] Faculty of Information Technology, Prince Al-Hussein Bin Abdullah-II for IT, Hashemite University, P.O.Box.330127, Zarqa, JORDAN.

[2] Department Of Computer Science and Information Technology, Institute of Technology and Management ,MD University, Gurgaon, Haryana, INDIA.

Summary
This paper the authors propose a new method of intelligent search called vague-search to find the most suitable match for the predicates to answer any imprecise query made by the database users. The method is based on the theory of vague sets introduced by Gau and Buehrer. A corresponding fuzzy method could be generated a special case of our proposed method it is also to be mentioned that the vague-search method could be easily incorporated in the existing commercial query languages of DBMS to serve the lay users better.

Keywords:
Vague set, fuzzy set, β-value of an interval, vague relation, α-vague-equality search, vague-proximity search, vague-search, vague-predicate, vague-query, α-fuzzy equality search, fuzzy-proximity search, fuzzy-search, fuzzy-predicate, fuzzy-query, truth-value of a predicate.

1. Introduction

The traditional databases are designed to store and deal with data which are crisp, deterministic and precise in nature. The databases of industries, institutes or of any organizations are created on the basis of precise attributes with crisp and atomic data. In today's business world, a very rapidly growing number of databases are on the web, and are available to the users, most of them being lay users, at any corner of the world. The application of database systems moves outside the realm of crisp mathematical world to the real world, because of the trend and need of openness, transparencies and scope of accessibility that needs to handle any type of query in natural languages too. For example, consider a STUDENTS database which is

STUDENTS (STUDENT_NAME, ROLL_NO, SEX, AGE, EYE_COLOUR, PHONE_NO, GPA).

Suppose that the all required integrity constraints imposed on this database are on the domains of some attributes, given by

dom(AGE) = [17,28], dom(EYE_COLOUR) = {black, blue, red, pink},

dom(SEX) = {M, F}, dom(GPA) = [0,5]

and also the following :

(i) The attribute-values for the attribute NAME will consist of alphabets, blank and dot characters only, not other.

(ii) An attribute-value for the attribute ROLL_NO must be numeric of eight digits in length.

Now consider a crisp query in a QL made by a System-Manager like below :

PROJECT (STUDENT_NAME), WHERE $19 \leq AGE \leq 25$ and $3.5 \leq GPA \leq 4.5$.

The answer will

be immediately available. But if there is a query posed in natural language (by a lay user) like below :

PROJECT (STUDENT_NAME) WHO ARE "bright" AND "young", then the existing standard query languages will fail to answer it.

This failure is because of the presence of imprecise constraints in the query predicate which can not be tackled due to the limitation of the grammar in standard query languages which work on crisp environment only. But this type of queries are very common in business world and in fact more frequent than grammatical-queries, because the users are not always expected to have knowledge of DBMS and the query language.

Consequently, there is a genuine necessity for the different large size organizations, specially for the industries, companies having world wide business, to develop such a system which should be able to answer the users queries posed in natural language, irrespective of the QLs and their grammar, without giving much botheration to the users. Most of these type of queries are not crisp in nature, and involve predicates with fuzzy (or rather vague) data, fuzzy/vague hedges (with concentration or dilation). Thus, this type of queries are not strictly confined within the domains always. The corresponding predicates are not hard as in crisp predicates. Some predicates are soft because of vague/ fuzzy nature and thus to answer a query a hard match is not always found from the databases by search, although the query is nice and

Manuscript received May 5, 2009
Manuscript revised May 20, 2009

very real, and should not be ignored or replaced according to the business policy of the industry. To deal with uncertainties in searching match for such queries, fuzzy logic and rather vague logic[9] will be the appropriate tool.

In this paper we propose a new type of searching techniques using vague set theory of Gua and Buehrer [9] to meet the predicates posed in natural language in order to answer imprecise queries of the users. Thus it is a kind of an intelligent search for match in order to answer imprecise queries of the lay users. We call this method by 'vague-search of match for predicates' or in short by 'vague-search of predicates'.

A vague search of predicates is basically composed of two types of search which are :-

(i) α-vague-equality search, and

(ii) vague-proximity search

vague-search is a search in which a combination of the above two search techniques, as and when applicable, in addition to crisp search, is applicable to find the most suitable match. Therefore, first of all we will introduce the methods of these above two searches, and then finally we will introduce the method of vague-search. Quite naturally, if there is no element of indeterministic-part all-through the computation, the proposed method of vague-search reduces to new method of 'fuzzy search'. Our method of "fuzzy search" is thus a combination of the following searches, in addition to crisp search :-

(i) α-Fuzzy-equality search (as a special case of α-vague-equality search), and

(ii) Fuzzy-proximity search (as a special case of vague-proximity search).

Our fuzzy method of search is different from the existing methods [4, 6, 7, 11,13, 18]. Our method, being an intelligent soft-computing method, will support the users to make and find the answers to their queries without iteratively refining them by trial and error which is really boring and sometimes it seriously effects the insert (mission and vision) of the organization, be it an industry, or a company or a hospital or a private academic institution etc. to list a few only out of many. Very often the innocent (having a lack of DBMS knowledge) users go on refining their queries in order to get an answer. The users are from different corner of the academic world or business world or any busy world. For databases to support imprecise queries, our intelligent system will produce answers that closely match the queries constraints, if does not exactly. This important issue of closeness can not be addressed with the crisp mathematics. That is why we have used the vague tools (and fuzzy tools).

2. theory of vague sets [9]

First of all we recollect a basic preliminaries of vague theory.

There are a number of generalizations [1,2,3,9,10] of Zadeh's fuzzy set theory [19] so far reported in the literature viz., i-v fuzzy theory, two-fold fuzzy theory, vague theory,

Intuitionist fuzzy theory, probabilistic fuzzy theory, L-fuzzy theory, etc.. the notion of vague theory recently introduced in IEEE by Gau and Buehrer [9] is of interest to us for this present work. For each such generalization, one (or more) extra edge is added with the fuzzy theory with specialized type of aim and objective. Thus, a number of higher order fuzzy sets are now in literatures and are being applied into the corresponding more specialized application domains.

While fuzzy sets are applicable to each of such application domains, higher order fuzzy sets can not, because of their specialization in character by birth. Application of higher order fuzzy sets make the solution-procedures more complex, but if the complexity on computation-time, computation-volume or memory-space are not the matter of concern then a better result could be achieved. Vague sets defined recently by Gue and Buehrer [9] have also an extra edge over fuzzy sets. Let U be a universe, say the collection of all students of Calcutta High School. Let A be a vague set of all "good-in-maths students" of the universe U, and B be a fuzzy set of all "good-in-maths students" of U. Suppose that an intellectual Manager M_1 proposes the membership value $\mu_B(x)$ for the element x in the fuzzy set B by his best intellectual capability. On the Contrary, another intellectual Manager M_2 proposes independently two membership values $t_A(x)$ and $f_A(x)$ for the same element in the vague set A by his best intellectual capability. They amount $t_A(x)$ is the true-membership value of x and $f_A(x)$ is the false-membership value of x in the vague set A . Both M_1 and M_2 being human agents have their limitation of perception, judgment, processing-capability with real life complex situations. In this case of fuzzy set B, the manager M_1 proposes the membership value $\mu_B(x)$ and proceed to his next computation. there is no higher order check for this membership value in general. In the later case, the manager M_2 proposes independently the membership values $t_A(x)$ and $f_A(x)$, and makes a check at this base-point it self by exploiting the constraint $tA(x) + fA(x) \leq 1$. if it is not honored, the manager has a scope of rethink, to reshuffle his judgment procesdure either on 'evidence against' or on 'evidence for' or on both. The two membership values are proposed independently, but they are mathematically not independent. They are mathematically constrained. This is the breaking philosophy in Gau and Buehrer's theory vague sets [9].

In this classical work[19], Zadeh proposed the theory of fuzzy sets. Since then it has been applied in wide varieties of fields like Computer Science, Management Science, Medical Science, Engineering problems etc. to list a few only.

Let U = { u_1,u_2………………..,u_n} be the universe of discourse. The membership function for fuzzy sets can have functional values from the closed interval [0,1]. A fuzzy set A in U is defined as the set of ordered pairs A = { (u, μ_A(u)) : u $\in$ U }, where μ_A(u) is the grade of membership of element u in the set A . The greater μ_A(u), the greater is the truth of the statement that 'the element u belongs to the set A'. With this philosophy, Prof. Zadah generalized the notation of crisp subset of classical set theory.
But Gau and Buehrer [9] pointed out that this single value combines the 'evidence for u' and the 'evidence against u'.It does not indicate the 'evidence for u' and the 'evidence against u' and it does not also indicate how much there is of each. Consequently, there is a genuine necessity of a model like vague sets, a kind of higher order fuzzy sets, which could be treated as a generalization of Zadeh's fuzzy sets[19].

Definition 2.1 A vague set (or in short VS) A in the universe of discourse U is characterized by two membership functions given by :-

(i) a truth membership function
$t_A : U \rightarrow [0,1]$**, and**

(ii) a false membership function
$f_A : U \rightarrow [0,1]$,

where $t_A(u)$ is a lower bound of the grade of membership of u derived from the 'evidence for u', and $f_A(u)$ is a lower bound on the negation of u derived from the 'evidence against u', and their total amount can exceed 1 i.e. $t_A(u) + f_A(u) \leq 1$.

Thus the grade of membership of u in the vague set A is bounded by a subinterval [$t_A(u)$, 1- $f_A(u)$] of [0,1].This indicates that if the actual grade of membership is $\mu(u)$, then $t_A(u) \leq \mu(u) \leq 1 - f_A(u)$.

The vague set A is written as A= {< u, [$t_A(u)$,$f_A(u)$]>:u $\in$ U}, where the interval [$t_A(u)$, 1-$f_A(u)$] is called the 'vague value' of u in A and is denoted by $V_A(u)$.
For example, consider an universe U = {DOG, CAT, RAT}. A vague set A of U could be A = { <DOG,[.7,.2]>, <CAT,[.3,.5]., <RAT,[.4,.6]> }.

It is worth to mention here that interval-values fuzzy sets (i-v fuzzy sets) [20] are not vague sets. In i-v fuzzy sets, an interval valued membership value is assigned to each element of the universe considering the 'evidence for u' only, without considering 'evidence against u'. in vague sets both are independently proposed by decision maker. This makes a major difference in the judgment about the grade of membership.

Definition 2.2 Zero Vague Set and Unit Vague Set.

A vague set A of a set U with $t_A(u)=0$ and $f_A(u)=1$ $\forall$ u $\in$ U is called the zero vague set of
U. A vague set A of a set U with $t_A(u)= 1$and $f_A(u)= 0$ $\forall$ u $\in$ U is called the unit vague set of U. The definition of α -vague set is also analogous.

Definition 2.3 A vague set A of a set U with $t_A(u)=\alpha$ and $f_A(u)=1-\alpha$ $\forall$ u $\in$ U is called the α –vague set of U, where $\alpha \in [0,1]$.

Definition 2.4 A vague number (VN) is a vague set of the set R of real numbers.
The theory of vague sets introduced by Gau and Buehrer [9] is an identical concept of the theory of intuitionistic fuzzy sets introduced by Atanassov [1,2], as pointed out by Bustince and Burillo in [3].

Definition 2.5 If A and B are two vague of the set E, then

$A \subset B$ iff $\forall x \in E$, [$t_A(x) \leq t_B(x)$ and $f_A(x) \geq f_B(x)$].

$B \supset A$. iff $A \subset B$

A = B iff $\forall$ x $\in$ E, [$t_A(x) = t_B(x)$ and $f_A(x)=f_B(x)$].

$\overline{A} = \{ \langle x, f_A(x), t_A(x) \rangle \mid x \in E \}$.

$A \cap B = \{ \langle x, \min(t_A(x),t_B(x)), \max(f_A(x), f_B(x)) \rangle \mid x \in E \}$.

$A \cup B = \{ \langle x, \max(t_A(x),t_B(x)), \min(f_A(x),f_B(x)) \rangle \mid x \in E \}$.

3. Vague Relation and their Properties

Fuzzy relation ([16],[21]) have a wide range of applications in different areas in Computer Science, in Management Science, in Banking and Finance, in Social Science etc. Our work this paper is based on theory o relations, specially vague relations. In this section we recollect the recent literature [14,15] on the nation of vague relations and their properties. First of all we mention the following notations on interval arithmetic which will be used in our work here subsequently.

Notations

Let I[0,1] denotes the family of all closed subintervals of [0,1]. If I_1=[a_1,b_1] and I_2[a_2,b_2] be two elements of I[0,1], we call $I_1 \geq I_2$ if $a_1 \geq a_2$ and $b_1 \geq b_2$. similarly we understand the relations $I_1 \leq I_2$ and $I_1 = I_2$. Cleary the the relation $I_1 \geq I_2$ does not necessarily imply that $I_1 \supseteq I_2$ and conversely. Also for any two unequal intervals I_1 and I_2, there is no necessity that either $I_1 \geq I_2$ or $I_1 \leq I_2$ will be true. the term 'imax' means the maximum of two intervals as imax(I_1,I_2)= [max(a_1,a_2),max(b_1,b_2)]. Similarly

defined is 'imin'. The concept of 'imax' and 'imin' could be extended to define 'isup' and 'iinf'of infinite number of elements of I[0,1].
It is obvious that L = { I[0,1], isup, iinf, $\leq$} is a lattice with universal bounds [0,0] and [1,1].

3.1 Vague Relation (VR)

Let X and Y be two universes. A vague elation (VR) denoted by $R(X \rightarrow Y)$ of the universe X with the universe Y is a VS of the Cartesian product $X \times Y$.
The true membership value $t_R(x,y)$ estimates the strength of the existence of the relation of R-type of the object x with the object y , whereas the false membership vlue $f_R(x,y)$ estimates the strength of the non-existence of the relation of R-type of the object x with the object y. The relation $R(X \rightarrow Y)$ could be in short denoted by the notation R , if there is no confusion.

Example :

Consider two universes X = {a,b} and Y={p,q,r}. Let R be a VR of the universe X with the universe Y proposed by an intelligent agent as shown by the following table :-

VR $R(X \rightarrow Y)$

$R(X \rightarrow Y)$	p	q	r
x	(.7,.2)	(.3,.5)	(.8,.2)
y	(.2,.4)	(.7,.3)	(.4,.4)

The proposed VR reveals the strength of vague relation of every pair $X \times Y$; For example, it reveals that the object y of the universe X has R-relation with the element p of Y with the following estimation :-

Strength of existence of the relation = .7
Strength of non-existence of the relation = .2

A relation $E(X \rightarrow Y)$ is called a complete Relation from the universe X to the universe Y if $V_E(x,y) = [1,1]$ $\forall (x,y) \in X \times Y$. A relation $\Phi(X \rightarrow Y)$ is called a Null Relation from the universe X to the universe Y if $V\Phi(x,y)=[0,0]$ $\forall (x,y) \in X \times Y$.

3.2 various Operation on VRs

For suitable application of vague relation, we must be aware of the different operations on them.
In this section we defined some operation on VRs.

Definition 3.2.1 Complement of a VR
Let $R(X \rightarrow Y)$ be a VR describing some relation R. Its complement denoted by $R^c(X \rightarrow Y)$ is the VR given by

$$V_{R^c}(x,y) = [f_R(x,y), 1-t_R(x,y)]$$

Definition 3.2.2 Union of two VRs
Let $R(X \rightarrow Y)$ and $S(X \rightarrow Y)$ be two VRs each from the universe X to the universe Y . The union of R and S is denoted by $R \cup S$ which is also a VR from X to Y, and is given by $V_{R \cup S}(x,y) = \text{imax}\{V_R(x,y), V_S(x,y)\}$.

Definition 3.2.3 Intersection of two VRs

Let $R(X \rightarrow Y)$ and $S(X \rightarrow Y)$ be two VRs each from the universe X to the universe Y. The intersection of R and S is denoted by $R \cap S$ which is also a VR from X to Y, and is given by

$$V_{R \cap S}(x,y) = \text{imin}\{V_R(x,y), V_S(x,y)\}.$$

The following proposition is straightforward.

Proposition 3.2.1
Let $R(X \rightarrow Y)$, $S(X \rightarrow Y)$ and $T(X \rightarrow Y)$ be three VRs each relating the universe X with the universe Y. Then

(i) $(R^c)^c = R$
(ii) $R \cup S = S \cup R$
(iii) $R \cap S = S \cap R$
(iv) $R \cup (S \cup T) = (R \cup S) \cup T$
(v) $R \cap (S \cap T) = (R \cap S) \cap T$
(vi) $R \cup (S \cap T) = (R \cup S) \cap (R \cup T)$
(vii) $R \cap (S \cup T) = (R \cap S) \cup (R \cap T)$
The laws of Excluded Middle do not hold here. i.e.
(viii) $R \cup R^c \neq E$ and $R \cap R^c \neq \Phi$

De Morgan's laws hold good in the operation with VRs.

Proposition 3.2.2

Let $R(X \rightarrow Y)$ and $S(X \rightarrow Y)$ be two VRs each relating the universe X with the universe Y.
Then the following results are true:-
(i) $(R \cup S)^c = R^c \cap S^c$
(ii) $(R \cap S)^c = R^c \cup S^c$

3.3 Composition of VRs

A vague set and a vague relation, under a suitable composite, could yield a new vague relation with an useful significance. Similarly two vague relation, under a suitable composition, could too yield a new vague relation with an useful significance. Composition of a relation is important for application, because of the reason that if a relation of a universe X with another universe Y is know and if a relation of the universe Y with a third universe Z is known then the relation of X with Z could be computed.

Definition 3.3.1 Composition of a VS and a VR

Let A be a VS of the universe X and R be a VR of the universe X with another universe Y. The composition of R with A, denoted by B = R $_O$ A, is a VS in Y given by

$V_{RoA}(y) = [\ \text{isup}_{x\in X}\ \{ t_A(x) \wedge t_R(x,y)\},\ \text{isup}_{x\in X}\ \{(1-f_A)(x) \wedge (1-f_R)(x,y)\}\ ]$.

Definition 3.3.2 Composition of two VRs

Let $R(X\to Y)$ and $S(Y\to Z)$ be two VRs. Then the composition relation B=R o S is a VR of X with Z given by

$V_{ROS}(x,z) = [\ \text{isup}\ y\in Y\ \{t_R(x,y)\wedge t_S(y,z)\ \},\ \text{isup}\ y\in Y\ \{(1-f_R)(x,y)\wedge(1-f_S)(y,z)\}]$.

This composition yields a vague –valued link between the objects x (of X) and z (of Z) through the elements y(of Y) .

Clearly R o S $\neq$ S o R.

Consider a vague relation $R(X\to X)$ of a universe X with itself . this type of relation we will call as a 'VR R on the universe X'.

Definition 3.3.3

A VR R on a universe X is said to be

(i) reflexive : if $\forall x \in X$, $V_R(x,x) = [1,1]$.

(ii) Symmetric : if $\forall x_1, x_2 \in X$, $V_R(x_2,x_1) = V_R(x_2,x_1)$.

Next we present the concept of inverse of a VR $R(X\to Y)$.

Definition 3.3.4

Let $R(X\to Y)$ be a VR relating X with Y . Then its inverse is denoted by $R^{-1}(Y\to X)$ which is a VR relating Y with X and is given by

$$V_{R-1}(y,x)= V_R(x,y) \quad \forall x \in X \text{ and } \forall y\in Y$$

It may be noticed that inverse and complement of a VR are two different concepts.

Proposition 3.3.1

Let $R(X\to Y)$ and $S(Y\to Z)$ be two VRs. Then the following are true:

(i) $(R^{-1})^{-1} = R$

(ii) $(R \text{ o } S)^{-1} = S^{-1} \text{ o } R^{-1}$

Definition 3.3.5

A VR R on a universe X is said to be a vague tolerance relation (VTR)or a vague proximity relation (VPR)on X if it is both reflexive and symmetric.

Example (of a VPR or VTR):-

Suppose that, in biotechnology experiment four potentially new strains of bacteria B_1,B_2,B_3,and B_4 have been detected in the area around an anaerobic corrosion pit on a new aluminum-lithium alloy used in the fuel tanks of a new experimental aircraft. In order to propose methods to eliminate the bio- corrosion caused by these bacteria , the four strains must be categorize first of all .One way to categorize them is to compare them to one another. In a pair –wise comparison, the following VPR is developed:-

R	B1	B2	B3	B4
B_1	[1,1]	[.3,.4]	[.8,.1]	[0,0]
B_2	[.3,.4]	[1,1]	[.6,.4]	[.2,.3]
B_3	[.8,.1]	[.6,.4]	[1,1]	[.8,.9]
B_4	[0,0]	[.2,.3]	[.8,.9]	[1,1]

The following proposition is straightforward

Proposition 3.3.2

If R_1 and R_2 be two VTRs on X, then

(i) R_1^{-1} is also a VTR on X.

(ii) $R_1 \bigcap R_2$ is a VTR on X.

(iii) $R_1 \bigcup R_2$ is a VTR on X.

In the next section ,let us present the classical relation model of Codd in brief for the sake of completeness.

4. The Classical Relational Model

A classical relational database ([8],[17]) consist of a collection of relations. A relation is a table of values where each row represents a collection of related data values. in a table , each row is called a tuple, a column header is called an attribute and the table as a whole is called the relation . A relation schema $R(A_1,A_2,.....A_n)$ consists of a relation name R and a list of attributes

$A_1,A_2,.....A_n$

The domain of an attribute A_i is denoted by $dom(A_i)$. An instance relation r of the relation schema $R(A_1,A_2,.....A_n)$ also denoted by r(R), is thus a set of tuples $t_1,t_2,....t_n$ where each t_i is an n-tuple of the form $t_i=\langle\ \nu_1, \nu_2,........ \nu_n)$, $\nu_i \in dom(A_i)$.

The ith value in tuple *t* corresponds to the attribute A_i and is denoted by $t[A_i]$. There are various restrictions on data in the form of constraints. Domain constraints specify that each value of an attribute A_i must be an atomic value from the domain dom(A_i).this includes restrictions on data types , on the range of values (if any),and on the format of data . Assume that null is not an element of any domain dom (A_i), the entity integrity constraint, which states that no primary key value can be null, is satisfied . the key constraint says that if $K \subseteq R$ is a super-key then for any two distinct tuples t_1 and t_2 in r(R) ,we have the constraint that $t_1[K] \neq t_2[K]$.

The referential integrity constraints are not something imposed on any individual relation in a database. It is specified between two relations, and is used to maintain the consistency among tuples of the two relations. If we want to study the logical design of a relational database, we see that the integrity constraints play an important role ([8],[17]).

5. A Note on Interval Mathematics

Vague set theory of Gau and Buehrer[9] and interval-valued fuzzy set theory of Zadeh [20] are two different concepts. These two theories should not be confused to be analogous. In interval –valued fuzzy set theory , the membership values are proposed on the philosophy of membership –property only, without considering the fact of non-membership . This is not the case in vague set theory . However ,while dealing with the mathematics of vague set theory ,the script theory of interval mathematics is sometimes useful.

In this section we recollect some basic notions of interval mathematics. For our purpose in this paper , we need to consider intervals of non- negative real numbers only.

Let I_1=[a,b] and I_2=[c,d] be two interval of non-negative real numbers. A point valued non-negative real number R also can be viewed , for the sake of arithmetic, as an interval [r,r].

5.1 Some Algebraic Operations

(i) Interval Addition $I_1 + I_2 = [a+c ,b+d]$
(ii) interval Subtraction: $I_1 - I_2 = [a-c,b-d]$
(iii) Interval Multiplication: $I_1 * I_2 = [ac,bd]$
(iv) interval Division: $I_1 \div I_2 = [a/c,b/d]$, when c, d $\neq 0$.
(v) Scalar Multiplication: $K.I_1 = [Ka,Kb]$.

5.2 Ranking of Intervals

Intervals are not ordered . Owing to this major weakness, there is no universal method of ranking a finite(or infinite)number of intervals . But in real life problems dealing with intervals we need to have some tactic to rank them in order to arrive at some conclusion. we will now present a method of ranking of intervals, which we shall use in our work here in subsequent section.

We consider a decision maker (or any intelligent agent like a company manager, a factory supervisor, an intelligent robot, an intelligent network, etc.)who makes a pre-choice of a decision-parameter $\beta \in [0,1]$. The intervals are to be ranked once the decision –parameter β is fixed. . But ranking may differ if the pre –choice β is renewed .

Definition 5.2.1 β –value of an interval

Let J=[a,b] be an interval . the β-value of the interval J is a non –negative real number J_β, given by $J_\beta = (1- \beta).a + \beta.b$.

Clearly, $0 \leq J_\beta \leq 1$, and $\beta = 0$ $J_\beta = a$ which signifies that the decision –maker is pessimistic, and also for β=1 $J_\beta = b$ which signifies that the decision maker is optimistic. For β=.5 it is the arithmetic-mean to be choosen usually for a moderate decision.

Comparison of two or more intervals we will do here on the basis of β-values of them. If the value of β is renewed, the comparison–results may change. The following definition will make it clear.

Definition 5.2.2 Comparing two intervals

Let J_1=[a,b] and J_2=[c,d] be two intervals. Then for a choosen $\beta \in [0,1]$, we define

(i) $J_1 < J_2$ if $(J_1)_\beta < (J_2)_\beta$.
(ii) $J_1 > J_2$ if $(J_1)_\beta > (J_2)_\beta$.
(iii) $J_1 = J_2$ if $(J_1)_\beta = (J_2)_\beta$.

(Note : The intervals J_1=[a,b] and J_2= [a,b] are strictly equal. For the other cases of the equality " “$J_1 = J_2$” , a further internal ranking could be done on the basis of their range i.e. interval-length. If range is more , we impose that the corresponding interval is greater).

We are now prepared to explain our proposed three methods of search in the next sections. First of all we introduce the two basic methods of search, which are α -vague –equality search and vague-proximity search, and finally we combine them to define the notion of vague-search.

6. a-vague-equality Search

Consider the STUDENT database as described in section -1. Consider a normal type of query like

PROJECT(STUDENT_NAME)

WHERE AGE = "approximately 20"

The standard SQL is unable to provide any answer to this query as the search for an exact match for the predicate will fail. The value"approximately20"is not a precise data. Any data of type "approximately x", "little more than x", "slightly less than x", “much greater than x" etc. are not precise or crisp, but they are vague numbers(VN)(see Definition 2.4) or fuzzy numbers which are special case of vague numbers. Denote any one of them , say the vague number: “approximately x" by the notation l(x). We know that the vague number is a VS of the set of real numbers. Clearly for every number a $\in$ dom (AGE) , There is a membership value $t_{I(x)}$ (a) proposing the degree of equality of this crisp number a with the quantity "approximately x", and a non-membership value $f_{I(x)}(a)$ proposing the degree of non equality . Thus ,in vague philosophy of Gau and Buehrer, every element of dom (AGE)satisfiesthe predicate AGE="approximately 20"up to certain extent and does not satisfy too, up to certain extent . but we will restrict ourselves to those members of dom(AGE) which are α – vague –equal, the concept of which we will define below. Any imprecise predicate of type AGE = "approximately 20" , or of type AGE ="young" (where the attribute value "young" is not a member of the dom(AGE)), is to be called by vague – predicate , and a query involving vague –predicate is called to be a vague-query. Consequently, as special cases of the vague terminologies, the concepts of fuzzy predicate ,fuzzy query and α -fuzzy –equality search are clear to us.

Definition 6.1

Consider a choice –parameter $\alpha \in [0,1]$. A member of a of dom (AGE) is said to be α -vague-equal to the quantity "approximate x" if a $\in I_\alpha$ (x) where is the α -cut of the vague number I(x). The degree or amount of this quality is measured bythe interval $m_{I(x)}(a) = [t_{I(x)}(a), 1- f_{I(x)}(a)]$.
Denote the collection of all such a-vague-equal members from dom (AGE)by the notation $AGE_\alpha(x)$, which is a subset of dom(AGE). If $AGE_\alpha(x)$ is not a null-set or singleton, then the members can be ranked by ranking their corresponding degrees of equality (using Definition 5.2.2).

Definition 6.2
Consider a choice value β $\in$ [0,1]. At β level of choice, for every element a of AGE_a (x), the truth –value $t(p_1,p_2)$ of the matching of the predicate p_1: given by AGE="approximately x" with the predicate p_2: AGE = a is equal to the β –value of the interval $m_{1(x)}(a)$.

7. Vague-proximity Search

The notion of α -vague-equality search as explained above is appropriate while there is an vague –predicate in the query involving VNs. But there could be a variety of vague predicates existing in a vague query, many of them may involve vague fuzzy hedges (including concentration/dilation) like "good", "very good", "excellent", "too much tall", "young", "not old", etc. In this section we present another type of search for finding out a suitable match to answer imprecise queries. In this search we will use the theory of vague –proximity relation [4,5]. We know that the vague –proximity relation on a universe U is a vague relation on U which is both vague –reflexive and vague –symmetric.
Consider the STUDENTS database as described in section -1 and a query like

PROJECT(STUDENT_NAME)

WHERE EYE-COLOR = "dark-brown".

The value /data "dark-brown" is not in the set dom(EYE-COLOR). There fore a crisp search will fail to answer this. The objective of this research work is to overcome this type of drawbacks of the classical SQL.
For this we notice that there may be one or more members of the set dom(EYE-COLOR) which may closely match the eye-color of “brown" or "dark brown".
Consider a new universe given by

$$W = \text{dom(EYE-COLOR)} \cup \{\text{dark-brown}\}.$$

Propose a vague –proximity relation R over W. Choose a decision –parameter α $\in$ [0,1]. We propose that search is to be made for the match e $\in$ dom(EYE-COLOR) such that

$t_R(\text{dark-brown},e) \geq \alpha$.

(It may be mentioned here that the condition $t_R(\text{dark-brown},e) \geq \alpha$ does also imply the condition $f_R(\text{dark-brown},e) \leq 1-\alpha$).

We say that e is a close match with "dark-brown" with the degree or amount of closeness being the interval $m_{\text{dark-brown}}(e)$ given by

$$m_{\text{dark-brown}}(e) = [t_R(\text{dark-brown},e), 1-f_R(\text{dark-brown},e)].$$

At β level of choice ,the truth –value $t(p_1,p_2)$ of the matching of the predicate p_1: given by EYE_COLOR ="dark-brown" with predicate p_2: AGE=e is equal to the β-value of the interval $m_{\text{dark-brown}}(e)$.

8. Vague-search

In this section we will now present the most generalized method of search called by vague –search. The vague search of matching is actually a combined concept of α – vague-equality search , vague-proximity search and crisp search.
For example, consider a query like
PROJECT(STUDENT_NAME)

WHERE (SEX="M",EYE-COLOR="dark-brown", AGE="approximately 20").
This is a vague query.
To answer such a query , matching is to be searched for the three predicates P1,P2 and P3 given by
P_1: SEX="M"
P_2: EYE_COLOR="dark-brown" and
P_3: AGE="approximately 20",
Where p_1 is crisp and P_2, P_3 are vague.
Clearly ,to answer this query the proposed vague search method is to be applied , because in addition to crisp search , both of a-vague-equality search and vague proximity search will be used to answer this query . the truth value of the matching of the conjunction P of P_1, P_2 and P_3 will be the product of the individual truth-values,(where it is needless to mention that for crisp match the truth –value will be exactly 1). There could be a multiple number of answers to this query , and the system will display all the results ordered or ranked according to the truth –values of p.
It is obvious that the vague –search technique for predicate-matching reduces to a new type of fuzzy-search technique as a special case.

9. Conclusion

In this paper, we have introduced a new method to answer imprecise queries of the lay users from the databases(details of the databases may not be known to the lay users). We have adopted vague set tool to solve the problem of searching an exact match or a close match (if an exact match is not available) of the predicates in order to extract the best answers (with appropriate ranks among them) of the users queries . As a special case ,the method reduces to a method of fuzzy search to answer imprecise queries . Our approach is in one sense domain independent if the query is of a vague nature. This supports the users to make any query in his natural language without the botheration of reframing his crisp queries repeatedly to ultimately gain some answer.
We claim that the method could be well incorporated in the existing commercial query languages so that the users of any level of knowledge can get some results to his queries. The system should have link to pre-formulated VNs corresponding to different domains of the databases, and also should have an intelligent on-line subsystem to create vague-proximity relations according to the vague –queries posed by the users from any corner of the world. This will be our next course of work which we will propose coding for implementation of this method with few examples. The search used to answer different queries suggested in ([4], [6], [7], [11], [13], [18])are not the same to our proposed method . Our vague –search as well as the corresponding fuzzy –search reported here are new proposals.

References

[1] Atanassov ,K., Intuitionistic Fuzzy Sets, Fuzzy Sets and Systems.20(1986):87-96.
[2] Atanassov, K., Intuitionistic Fuzzy Sets : Theory and Applications, physica-Verlag(2000), New-York.
[3] Bustince, H. and Burillo, P.,Vague sets are intuitionistic fuzzy sets, fuzzy sets and systems 79 (1996) 403-405
[4] Bosc P.,Lietard I and Pivert O,"Evaluation of flexible queries: The quantified statement case".Technologies for constructing Intellegent Systems I, physica-Verlag Heidelberg New Yourk,pp 337-350 (2002).
[5] Chen, Shyi-ming, Analyzing fuzzy systems reliability using vague set theory,Int. Jou. Of Applied Sc. & Engg. Vol.1(2003)82-88.
[6] Chiang D., Chow L.R. and Heien N"fuzzy information in extended fuzzy relational databases", fuzzy sets and systems 92, pp.1-10. (1997)
[7] Cohen,W., Integration of heterogeneous databases without common domains using queries based on textual similarity,proc.OF SIGMOND,June 1998,pages 201-212.
[8] Elmasri,R and Navathe,S.B., Fundementals of database Systems ,PearsonEducation Asia, Delhi,2002.
[9] Gau W.L.and Buehrer, D,J, Vague sets , IEEE Transaction on systems, Man and Cybernetics,Vol.23 (1993)610-614.
[10] Goguen,J.A., L-fuzzy sets , Jou.maths.anal.Appl.19(1967)145-174.
[11] Haveliwala,T.,Gionis,A, Klein,D,and Indyk,P,Evaluatingstrategies for similarity search on the web , proceeding of WWW, Hawai, USA, may 2002.
[12] Hirota,K., Concepts of probabilistic sets,fuzzy sets and systems. 5(1)(1981)31-46.
[13] Kaushik S,and Nanda,H., Web based access of relational databases using fuzzy Natural Language Queries, International conference on cognitive Systems,Delhi,India, (1999).
[14] Khan,H,Ahmd,m., and Biswas,R.,on Vague groups ,accepted for puplication in inter.Jour.of Computational Cognition.
[15] Khan,H Ahmd,M., and Biswas,R., Vague relations ,accepted for publiucation in Inter.Jour. of Computational Cognition.
[16] Klir G.J.and YuanB .,fuzzy sets and fuzzy logic: Theory and Applications,Printice Hall, Inc.Englewood Cliffs, (2001)N.J., USA.
[17] Silberschatz ,A., Korth,H.F., and Sundarshan,S., Database System Concepts(3rd edition), The Mc grow-hill Companies (New Delhi) (1997)
[18] Yang Q.,Zhang W., Liu C., Wu J., Nakajima H., and Rishe N.D.EffecientProcessing of nested Fuzzy Database, IEEE Trans. On Knowledge and Data Eng., Vol.13,No.6,pp 884-901,Nov./Dec 2001.
[19] Zadeh,L.A., Fuzzy sets ,Inform and Control,8 (1965)338-356.
[20] Zadeh,L,A., The Concept of a linguistic variable and its application to a proximate reasoning I, Inform and control. 8(1975)199-249.
[21] Zimmerman J.,Fuzzy set theory And Its Applications, Kluwer Academic Publishers,2001,Norwell,Massachusetts,U.S.A.