A LANGUAGE FOR FUZZY STATISTICAL DATABASE

S.Guglani[1] and C.P. Katti[2]

School of Computer and Systems Sciences,
Jawahar Lal Nehru University, Delhi, India
[1]seema_guglani@yahoo.co.in
[2]cpkatti@yahoo.com

ABSTRACT

Fuzzy statistical database is a database used for fuzzy statistical analysis purpose. A fuzzy statistical table is a tabular representation of fuzzy statistics and is a useful data structure for fuzzy statistical database. Primitive fuzzy statistical tables are a building block of fuzzy statistical table. In this paper we defined the fuzzy statistical join operator in the framework of fuzzy statistical database. The fuzzy statistical dependency preservation property will be discussed for the fuzzy statistical join. We also propose a set of fuzzy statistical table manipulation operators for arbitrary fuzzy statistical tables and discuss an implementation for them. These findings offer important insights into the retrievability of information from a fuzzy statistical database.

KEYWORDS

statistical database, fuzzy statistical database, fuzzy statistical equality, fuzzy statistical dependency

1. INTRODUCTION

Many researchers have explored the fundamentals of statistical database ([2], [4], [6-9],[12]).Most of the existing statistical database models are designed under the assumptions that the data/information stored is précised and queries are crisp. In fact, these assumptions are often not valid for many of the next generation database systems since they may involve information with uncertainty. In general, data /information in databases may be uncertain for the following reasons:

1. A decision in much knowledge-intensive application usually involves various forms of uncertainty.
2. Integrating data from various sources is not usually a crisp process, while unifying various heterogeneous data into an integrated form, due to semantic differences (and other reasons), sometimes forcing data to be completely crisp may result in falsity and useless information.
3. Information in some nontraditional applications is inherently both complex and uncertain i.e. representing subjective opinions and judgments concerning medical diagnosis, economic forecasting or personal evaluation.
4. In natural languages, numerous linguistic terms with modifiers (e.g. very, more or less etc.) and quantifiers (e.g. many, few, most etc.) are used when conveying vague information.

Handling uncertainty in data bases were first proposed on relational based database models. The last two decades have witnessed a blossoming of researches on this topic ([3],[5],[10],[11],[13],[14],[20-21],[22-23],[24-27]).

DOI: 10.5121/ijdms.2013.5106

Uncertainity in statistical database([4],[9],[15],[16]) was introduced by Seema[28].As a result fuzzy statistical database was developed. Fuzzy statistical tables are used to represent fuzzy statistics in fuzzy statistical database. The use of fuzzy statistical tables is not restricted to outputting formatting; they are maintained for bookkeeping, comparison and evaluated over a time span. So, we need a data manipulation language for fuzzy statistical tables.In this paper, we propose the set of operators to manipulate fuzzy statistical tables. Join and projection operations are also defined. These set of operators have the capability to express arbitrary queries involving fuzzy statistical tables.

The paper is organized as follows. In section 2, we introduce the preliminaries which include fuzzy statistical database, fuzzy primitive statistical table and fuzzy statistical equality. In section 3, fuzzy statistical operations are defined. Finally in section 4, implementation of fuzzy statistical operations is discussed.

2. PRELIMINARIES

2.1. Fuzzy Statistical Database

Fuzzy Statistical Database [28] is a statistical database which allows imprecise or vague statistical data. Such type of database is quite useful when the information available is subjective and imprecise. The imperfect information is incorporated in the fuzzy statistical database in the form of fuzzy attributes and fuzzy statistics. It is important that the fuzzy statistical database which incorporates imprecision, be able to appropriately propagate the level of uncertainty associated with the data to the level of uncertainty associated with answers or conclusions based on data .Fuzzy statistics is organized in a fuzzy statistical database as fuzzy statistical tables—two-dimensional matrices made up of row header and column header where row header and column header are structured in the form of an ordered set of trees called fuzzy row or fuzzy column attribute forests. Each cell in a fuzzy statistical table has an associated set of fuzzy row or fuzzy column attributes. The set of fuzzy row and fuzzy column attributes of a cell forms a path from the root to a leaf in a fuzzy row and fuzzy column attribute tree. Each cell in a fuzzy statistical table is labeled by an attribute called cell attribute. A fuzzy statistical table scheme $FS(F_r, F_c, C)$ is a three tuple where F_r denote the fuzzy row attribute forest, F_c denote the fuzzy category attribute forest. C is the fuzzy statistics, represented with an additional two dimensional array of cells for μ denoting the membership degree of fuzzy statistics. A parenthesized expression to specify a fuzzy attribute tree which is a preorder enumeration of the tree (i.e. first the root then the subtrees from left to right) is used. Let C be the fuzzy statistics and $RA_1, ..., RA_n$ be fuzzy row attributes with their appropriate universes $RU_1, ..., RU_n$ respectively which forms a path from the root to a leaf in fuzzy row attribute tree for accessing fuzzy statistic C and $CB_1, ... CB_m$ be fuzzy column attributes with their appropriate universes $CU_1, ..., CU_m$ which forms a path from root to a leaf in a fuzzy column attribute tree for accessing fuzzy statistic C. Then fuzzy statistical table FS is defined as

$$FS(RA_1(RA_2(RA_3 ...... RA_n)),(CB_1(CB_2(CB_3 .....CB_m)),(C))$$

A fuzzy statistical table instance is a collection of cell instances structured as specified by the fuzzy statistical table scheme. A cell instance consists of value of its fuzzy row and fuzzy column category attribute and a value for its fuzzy statistic along with its membership degree. Depending upon the complexity of domain of fuzzy row and fuzzy column attributes of fuzzy statistical table, it can be classified into two categories-

(a) Type-1 Fuzzy Statistical Table
(b) Type-2 Fuzzy Statistical Table

Type-1 Fuzzy Statistical Table[28]. If the fuzzy row and fuzzy column attributes of fuzzy statistical table are of type-1 then it is called type-1 fuzzy statistical table.

Example 1. Consider a fuzzy statistical table scheme

2012COUNT(State(Sex(Exp,Sal)),(Incometax),(Count))

of highly salaried, highly paying incometax and highly experienced people in a sample of a population. The fuzzy statistic being measured is the fuzzy count[28] represented by cell attribute count where F_r is fuzzy row attribute forest consisting of single tree with fuzzy attributes State, Sex, Experience and Salary. Experience and Salary are denoted by State, Sex, Exp and Sal respectively, F_c is fuzzy column attribute forest consisting of single tree with fuzzy attribute Incometax.Count is the fuzzy count of male and female people in a state who are highly experienced and are paying high incometax or having high salary in a sample of a population. Table1 shows an instance of fuzzy statistics table 2012COUNT. In 2012COUNT there are 128 instances for the cell attribute Count with corresponding 128 instances characterizing their fuzziness. Suppose the Universe of discourse for the Exp, U_{Exp} is the set of positive integers in the range 0-30, Universe of discourse for Sal ,U_{Sal} is the set of integers in the range 10,000-100,000, Universe of discourse for Incometax, $U_{Incomeax}$ is the set of integers in the range 0-10,000, Universe of discourse for State is {Delhi, Bombay} , Universe of discourse for Sex is {M,F}.Here domain of State and Sex are crisp sets whereas the domain of Experience, Incometax and Salary are fuzzy sets High-Exp,High-Sal and High-Incometax in their appropriate universes. i.e.

$dom(State) = \{Delhi, Bombay\}$
$dom(Sex) = \{M, F\}$
$dom(Exp) =$ High-Exp
$dom(Sal) =$ High-Sal
$dom(Incometax) =$ High-Incometax

The membership function μ_{HX}, μ_{HS} and μ_{HI} of the fuzzy sets High-Exp, High-Sal and High-Incometax, are as given below:

For $x \in U_{Exp}$,
$\mu_{HX}(x) = (1 + |x - 10|/4)^{-1}$ for $x \leq 10$
$= 1$ for $x > 10$
For $s \in U_{Sal}$,
$\mu_{HS}(s) = (1 + |s - 60{,}000|/20{,}000)^{-1}$ for $s \leq 60{,}000$
$= 1$ for $s > 60{,}000$
For $y \in U_{Incometax}$,
$\mu_{HI}(y) = (1 + |y - 5{,}000|/1{,}000)^{-1}$ for $y \leq 5{,}000$
$= 1$ for $y > 5{,}000$

Type-2 Fuzzy Statistical Table[28]. If the fuzzy row and fuzzy column attributes of fuzzy statistical table are of type-2 then it is called type-2 fuzzy statistical table.

Example 2. Consider a type-2 fuzzy statistical table scheme

FS1(State(Sex(Exp,Sal)),(Incometax),(Count1))

in a sample of a population shown in table 2. As in example 1, the Universe of discourse for the Exp, U_{Exp} is the set of positive integers in the range 0-30, Universe of discourse for Sal, U_{Sal} is the set of integers in the range 10,000-100,000, Universe of discourse for Incometax, $U_{Incomeax}$ is the set of integers in the range 0-10,000, Universe of discourse for State is {Delhi, Bombay} , Universe of discourse for Sex is {M, F}. Domain of State and Sex are crisp sets whereas the domain of Experience, Incometax and Salary are set of fuzzy sets in their respective universes .i.e.

$dom(Exp) = \text{set of fuzzy sets in } U_{Exp}$ ={Little,Mod,10,15-20}

$dom(Sal) = \text{set of fuzzy sets in } U_{Sal}$ ={30,000,High,Low,40,000−60,000}

$dom(Incometax) - \text{set of fuzzy sets in } U_{Incomeax} - \{High, Low, 6{,}000, 4{,}000 - 7{,}000\}$

The membership functions of the fuzzy set descriptors High, Low, Little and Mod is domain dependent and are as given below.

For $x \in U_{Exp}$.

$\mu_{Mod}(x) = (1 + |x - 8|)^{\ 1}$ for $x > 1$

= 0 otherwise

$\mu_{Little}(x) = (1 + 12x)^{-1}$ for $x > 0$

= 0 otherwise

For $y \in U_{Sal}$.

$\mu_{High}(y) = (1 + |y - 60{,}000|/20{,}000)^{-1}$ for $y \le 60{,}000$

$= 1$ for $y > 60{,}000$

$\mu_{Low}(y) = 1 - \mu_{High}(y)$

For $y \in U_{Incometax}$.

$\mu_{High}(y) = (1 + |y - 5{,}000|/1{,}000)^{-1}$ for $y < 5{,}000$

$= 1$ for $y > 5{,}000$

$\mu_{Low}(y) = 1 \quad \mu_{High}(y)$

The fuzzy statistics Count1 is the fuzzy count of male and female people in a state who are experienced or salaried and are paying incometax in a sample of a population.

2.2. Fuzzy Primitive Statistical Table

A fuzzy statistical table $FS(F_r, F_c, C)$ is a fuzzy primitive statistics table if $|F_r| = 1, |F_c| = 1$ and each tree in F_r and F_c has exactly one leaf .The fuzzy statistical table shown in table 2 consists of two fuzzy primitive statistics table as $|F_r| = 1, |F_c| = 1$ and the tree in F_r has two leaves. The instance of two fuzzy primitive statistics table of above example is shown in table 3 and table 4 respectively.

2.3. Fuzzy Statistical Dependency

Fuzzy integrity constraints are introduced in fuzzy statistical database by defining the dependency between its attributes. Knowledge of dependency between the attributes of fuzzy statistical database allows to obtain a correct logical model of fuzzy statistical database. Consider a fuzzy statistical table scheme (F_r, F_c, C) . The cell *C* in FS is dependent upon the attributes in row header and column header of fuzzy statistical table in a sense that if the instances of attributes in row header and column header are more or less equal then the corresponding instances of cell will also be more or less equal. Suppose for accessing the instance c of cell *C* in fuzzy statistical table FS, $X = (RA_1, RA_2, \ldots . RA_n)$ be fuzzy row attributes which forms a path from the root to a leaf in a fuzzy row attribute tree and $Y = (CB_1, CB_2, \ldots CB_m)$ be fuzzy column attributes which forms a path from the root to a leaf in a fuzzy column attribute tree with row instances $ra_1, ra_2, \ldots ra_n$ and column instances cb_1, cb_2,.... cb_m respectively. Also for the same fuzzy row and fuzzy column attributes X and Y, let $r'a_1, r'a_2, \ldots r'a_n$ be the instances of fuzzy row attributes and $c'b_1, c'b_2, \ldots c'b_m$ be the instances of fuzzy column attributes for accessing the instance $c'$of cell C then if the sets $\{ra_1, ra_2, \ldots ra_n\}$ and $\{cb_1, cb_2, \ldots cb_m\}$ are more or less equal to the sets $\{r'a_1, r'a_2, \ldots r'a_n\}$ and $\{c'b_1, c'b_2, \ldots c'b_m\}$ respectively then the corresponding cell instance c and c' would also be more or less equal. Seema[29] defined the fuzzy statistical dependency as follows:

Let X and Y be the set of fuzzy attributes in fuzzy statistical table in row header and column header respectively for accessing cell C, then the fuzzy statistical dependency $XY \rightsquigarrow C$ holds in fuzzy statistical table if and only if

$\forall t_1, t_2 \in X \times Y \times C$ such that $t_1' = t_1[XY]$, $t_2' = t_2[XY], t_1'' = t_1[C], t_2'' = t_2[C]$ we have

$$E_{XYC}(t_1, t_2) \wedge E_{XY}(t_1', t_2') \leq E_C(t_1'', t_2'')$$

where $E_A(a, b)$ denote the fuzzy statistical equality[29] for attribute A in fuzzy statistical table with instances a and b.

3. Fuzzy Statistical Table Operations

During the preliminary stage of data analysis for certain operations, often the statistician does not need to use the entire data set. Instead, to enhance responsiveness, the statistician may base his preliminary analysis on fuzzy primitive statistical tables which are building blocks of fuzzy statistical table. This motivates us to design the manipulation language by first defining operations to construct fuzzy primitive statistical table from fuzzy statistical table and vice-versa. Then extending the language to deal with arbitrary fuzzy statistical tables. Consider two fuzzy statistical table schemas

$$FS1\Big(\text{state}\big(\text{sex}(\text{exp}, \text{salary})\big), (Incometax), (Count1)\Big)$$

and

$$FS2(\text{state}(\text{sex}(\text{age})), (Incometax), (Count2))$$

A typical query on fuzzy statistical table may be to obtain a fuzzy statistics of state where only the experience is required or to obtain a fuzzy statistics of state where experience, salary and age are required. Such queries motivates us to define projection and join operations in fuzzy statistical environment using the physical organization technique [28] for fuzzy statistical table in which fuzzy row forest F_r is put into ordered tree TR by making the root nodes in F_r as immediate descendents of a dummy attribute θ_r and fuzzy column attribute forest F_c is put into ordered tree TC by making the root nodes in F_c as immediate descendents of dummy attribute θ_c.

3.1. Fuzzy Statistical Join Operator

Definition. Let FS1(F_{r1},F_{c1}, $Count1$) and FS2(F_{r2},F_{c2}, $Count2$) be two fuzzy statistical tables. Let X1 be the immediate descendent of θ_{r1} in TR1, X2 be the immediate descendent of θ_{r2} in TR2, Q1 be the immediate descendent of θ_{c1} in TC1, Q2 be the immediate descendent of θ_{c2} in TC2, Y1 be the extreme left leaf node of tree with root X1, Y2 be the extreme left leaf node of tree with root X2, R1 be the extreme left leaf node of tree with root Q1, R2 be the extreme left leaf node of tree with root Q2. Fuzzy row attribute forest F_{r1} and fuzzy column attribute forest F_{c1} of fuzzy statistical FS1 are said to be fuzzy statistical join compatible with fuzzy row attribute forest F_{r2} and fuzzy column attribute forest F_{c2} of fuzzy statistical FS2 if either $TR1$ and TR2 or TC1 and TC2 are same or if P1 , the path of attributes from X1 to predecessor of Y1 in TR1 is same as the path of attributes from X2 to predecessor of Y2 in TR2 or if P2 , the path of attributes from Q1 to predecessor of R1 in $TC1$ is same as the path of attributes from Q2 to predecessor of R2 in TC2 .

Let FS1(F_{r1},F_{c1}, $Count1$) and FS2(F_{r2},F_{c2}, $Count2$) be two fuzzy statistical tables which are fuzzy statistical join compatible, then the fuzzy statistical join operation denoted by $JOIN_{FS}(FS1,FS2)$ produces a fuzzy statistical table $FS(F_r,F_c,C)$ defined as:

(i) If TR1 and TR2 are same in FS1 and FS2, then F_r of FS is F_{r1} and there are two cases:
  (a) if P2, the path is same in TC1 and TC2 ,then F_c is formed by linking the leaf node R2 to predecessor of R1 in $TC1$ in extreme right direction .This process is repeated for all leaf nodes which are next to R2 and depending upon the root to leaf path instances in TR and TC , fuzzy statistics are taken care of. For example, if number of leaves in an instance of TC1 is n and number of leaves in an instance of TR1 is m, number of leaves in an instance of $TC2$ is r ,fuzzy statistics C will be represented by two dimensional array (n+r) x m with an additional two dimensional array (n+r) x m of cells for μ denoting the membership degree of fuzzy statistics. Corresponding to first root to leaf path instances of TR read only those fuzzy statistics corresponding to TC1 and write them into the fuzzy statistics array C then corresponding to first root to leaf path instances of TR read only those fuzzy statistics corresponding to TC2 and write them into the fuzzy statistics array C in continuation with the previous one. Simultaneously additional two dimensional array (n+r) x m of cells for μ is also read and written. This process is repeated for each root to leaf instance of TR.
  (b) otherwise $F_c = F_{c1} \cup F_{c2}$ and $C = Count1 \cup Count2$, the n^{th} row of the fuzzy statistics of FS2 is appended to the n^{th} row of that of FS1. $\cup$ denotes concatenation of ordered sets and C is an ordered fuzzy statistics set such that for each cell x in FS , if x is in FS1 then the fuzzy statistics of x is in Count1 otherwise it is in Count2.

(ii) If TC1 and TC2 are same in FS1 and FS2, then F_c of FS is F_{c1} and there are two cases:
  (a) if P1, the path is same in TR1 and TR2 ,then F_r is formed by linking the leaf node Y2 to predecessor of Y1 in $TR1$ in extreme right direction .This process is repeated for all leaf nodes which are next to Y2 and depending upon the root to leaf path instance in TR and TC, fuzzy statistics are taken care of. For example, if number of leaves in an instance of TC1 is n and number of leaves in an instance of TR1 is m, number of leaves in an instance of $TR2$ is r ,fuzzy statistics C will be represented by two dimensional array (m+r) x n with an additional two dimensional array (m+r) x n of cells for μ denoting the membership degree of fuzzy statistics. Read from FS1 and write into fuzzy statistics C the entries corresponding to the first instance of node which is predecessor to Y1 then read from FS2 and further write into the fuzzy statistics C the entries corresponding to the first instance of node which is predecessor to Y2.Repeat this for all other instances of node which are predecessor to Y1 and Y2. Simultaneously additional two dimensional array (m+r) x n of cells for μ is also read and written.
  (b) otherwise $F_r = F_{r1} \cup F_{r2}$ and $C = Count1 \cup Count2$, the fuzzy statistics of FS2 is appended to that of FS1. $\cup$ denotes concatenation of ordered sets and C is an ordered fuzzy statistics set such that for each cell x in FS , if x is in FS1 then the fuzzy statistics of x is in Count1 otherwise it is in Count2.

If neither $TR1$ and TR2 nor TC1 and TC2 nor P1, the path in TR1 and TR2 nor P2, the path in TC1 and TC2 are same then instead of join operation we can apply concatenation operation in fuzzy statistical tables.

3.2. Concatenation of Fuzzy Statistical Table

It is a binary operator which concatenates two fuzzy statistical tables having a common dimension of fuzzy row attribute forest and fuzzy column attribute forest. Let FS1 and FS2 be two fuzzy statistical tables $FS1(F_{r1},F_{c1},A_1)$ and $FS2(F_{r2},F_{c2},A_2)$ respectively. Then the operation

$$Conc_c(FS1,FS2)$$

produces a fuzzy statistical table FS with scheme $FS(F_r,F_c,A)$ where

$$F_c = F_{c1} \cup F_{c2}$$
$$A = A_1 \cup A_2$$

$\cup$ denotes concatenation of ordered sets and A is an ordered fuzzy statistics set such that for each cell x in FS , if x is in FS1 then the fuzzy statistics of x is in A_1 , otherwise it is in A_2. . The subscript of operation Conc can only be R or C denoting row or column concatenation respectively.

Example 3. Consider two fuzzy statistical table FS1 and FS2 whose scheme are given by
$FS1(\text{state}(\text{sex}(\text{exp}, \text{salary})), (Incometax), (Count1))$and
$FS2(\text{state}(\text{sex}(\text{exp}, \text{salary})), (Reward), (Count2))$ respectively
The column concatenation of $FS1$ and $FS2$, $Conc_c(FS1,FS2)$ produces FS given by $FS((state(sex(exp, salary)), (Incometax, Reward), (Count1, Count2))$

Theorem. The fuzzy statistical join operator in a fuzzy statistical database preserves fuzzy statistical dependency.

Proof. Suppose FS1 and FS2 be two fuzzy statistical table with schema $FS1(F_{r1},F_{c1},C1)$ and $FS2(F_{r2},F_{c2},C2)$.Let $X1 = R1A_1, R1A_2, \ldots .R1A_n$ and $Y1 = C1B_1, C1B_2, \ldots C1B_m$ be the set of fuzzy attributes occurring in the path from root to leaf in F_{r1} and F_{c1}respectively for accessing the cell attribute C1. Let $r1a_1, r1a_2, .. r1a_n$ be the instance of attributes $R1A_1, R1A_2, \ldots .R1A_n$ and $c1b_1, c1b_2, \ldots . c1b_m$ be the instance of attributes $C1B_1, C1B_2, \ldots C1B_m$ respectively for accessing the cell instance $c1_{kl1}$ of C1. Similarly, let $r'1a_1, r'1a_2, \ldots . r'1a_n$ be the instances of fuzzy row attributes $R1A_1, R1A_2, \ldots .R1A_n$ and $c'1b_1, c'1b_2, \ldots c'1b_m$ be the instances of fuzzy column attributes $C1B_1, C1B_2, \ldots C1B_m$ respectively for accessing the cell instance $c1_{kl2}$ of C1. Let $X2 = R2A_1, R2A_2, \ldots .R2A_n$ and $Y2 = C2B_1, C2B_2, \ldots C2B_m$ be the set of fuzzy attributes occurring in the path from root to leaf in F_{r2} and F_{c2} respectively for accessing the cell attribute C2. Suppose $r2a_1, r2a_2, .. r2a_n$ be the instance of attributes $R2A_1, R2A_2, \ldots .R2A_n$ and $c2b_1, c2b_2, \ldots . c2b_m$ be the instance of attributes $C2B_1, C2B_2, \ldots C2B_m$ respectively for accessing the cell instance $c2_{kl1}$ of C2.Similarly, let $r'2a_1, r'2a_2, \ldots . r'2a_n$ be the instances of fuzzy row attributes $R2A_1, R2A_2, \ldots .R2A_n$ and $c'2b_1, c'2b_2, \ldots c'2b_m$ be the instances of fuzzy column attributes $C2B_1, C2B_2, \ldots C2B_m$ respectively for accessing the cell instance $c2_{kl2}$ of C2. Since the fuzzy statistical dependency $X1Y1 \rightsquigarrow C1$ and $X2Y2 \rightsquigarrow C2$ holds in the fuzzy statistical table FS1 and FS2 respectively therefore
$\forall\, t1_1 = (r1a_1, \ldots r1a_i, .. r1a_n, c1b_1, \ldots c1b_m, c1_{kl1})$ and
$t1_2 = (r'1a_1, \ldots, r'1a_i, .. r'1a_n, c'1b_1, \ldots c'1b_m, c1_{kl2})$ in $X1 \times Y1 \times C1$, $t1'_1 = t1_1[X1Y1]$, $t1'_2 = t1_2[X1Y1]$, $t1''_1 = t1_1[C1]$, $t1''_2 = t1_2[C1]$ we have

$$E_{X1Y1C1}(t1_1,t1_2) \wedge E_{X1Y1}(t1'_1,t1'_2) \le E_{C1}(t1''_1,t1''_2)$$

and $\forall\, t2_1 = (r2a_1, r2a_2, .. r2a_n, c2b_1, \dots c2b_m, c2_{kl1})$ and $t2_2 = (r'2a_1, r'2a_2, .. r'2a_n, c'2b_1, \dots c'2b_m, c2_{kl2})$ in $X2 \times Y2 \times C2$, $t2'_1 = t2_1[X2Y2]$, $t2'_2 = t2_2[X2Y2]$, $t2''_1 = t2_1[C1]$, $t2''_2 = t2_2[C2]$ we have

$$E_{X2Y2C2}(t2_1,t2_2) \wedge E_{X2Y2}(t2'_1,t2'_2) \le E_{C2}(t2''_1,t2''_2)$$

Let FS1 and FS2 be fuzzy statistical join compatible and $FS(F_r, F_c, C)$ be the fuzzy statistical joined table .Suppose TR1 and TR2 are same in FS1 and FS2 and the path P2 is same in TC1 and TC2, then for accessing the cell attribute C, let $X = R1A_1, \dots R1A_n$ and $Y = C1B_1, .. C1B_{m-1}, C2B_m$ be the set of fuzzy attributes occurring in the path from root to leaf in F_r and F_c respectively. Let $r1a_1, r1a_2, .. r1a_n$ be the instance of attributes $R1A_1, R1A_2, \dots . R1A_n$ and $c1b_1, c1b_2, \dots, c1b_{m-1}, c2b_m$ be the instance of attributes $C1B_1, C1B_2, \dots, C1B_{m-1}, C2B_m$ respectively for accessing the cell instance $c2_{kl1}$ of C. Similarly, let $r'1a_1, r'1a_2, \dots . r'1a_n$ be the instances of fuzzy row attributes $R1A_1, R1A_2, \dots . R1A_n$ and $c'1b_1, c'1b_2, \dots c'1b_{m-1}, c'2b_m$ be the instances of fuzzy column attributes $C1B_1, C1B_2, \dots C2B_{m-1}, C2B_m$ respectively for accessing the cell instance $c2_{kl2}$ of C.

$\forall\, t_1 = (r1a_1, \dots r1a_i, .. r1a_n, c1b_1, \dots c2b_m, c2_{kl1})$ and
$t_2 = (r'1a_1, \dots, r'1a_i, .. r'1a_n, c'1b_1, \dots c'2b_m, c2_{kl2})$ in $X \times Y \times C$, $t'_1 = t_1[XY]$,
$t'_2 = t_2[XY]$, $t''_1 = t_1[C]$, $t''_2 = t_2[C]$

$$E_{XYC}(t_1,t_2)$$
$$= E_{R1A_1 \dots R1A_n, C1B_1, \dots C2B_m, C}\big((r1a_1, .., .r1a_n, c1b_1, . c2b_m, c2_{kl1}), (r'1a_1, ., r'1a_i, .. r'1a_n, c'1b_1, . c'2b_m, c2_{kl2})\big)$$
$$= E_{R1A_1}(r1a_1, r'1a_1) \wedge .. \wedge E_{R1A_n}(r1a_n, r'1a_n) \wedge E_{C1B_1}(c1b_1, c'1b_1) \wedge \dots \wedge E_{C2B_m}(c2b_m, c'2b_m)$$
$$\wedge E_C(c2_{kl1}, c2_{kl2})$$

Since $E_{R1A_i}(r1a_i, r'1a_i) = E_{R2A_i}(r2a_i, r'2a_i)$ for i=1…..n

and $E_{C1B_i}(c1b_i, c'1b_i) = E_{C2B_i}(c2b_i, c'2b_i)$ for i=1….m-1

therefore $E_{XYC}(t_1,t_2) = E_{X2Y2C2}(t2_1,t2_2)$, $E_{XY}(t'_1,t'_2) = E_{X2Y2}(t2'_1,t2'_2)$ and $E_C(t''_1,t''_2) = E_{C2}(t2''_1,t2''_2)$

Hence

$$E_{XYC}(t_1,t_2) \wedge E_{XY}(t'_1,t'_2) \le E_C(t''_1,t''_2)$$

So, fuzzy statistical dependency $XY \rightsquigarrow C$ holds in the fuzzy statistical table FS. Similarly we can show for other cases of fuzzy statistical join compatibility condition.

Hence, fuzzy statistical join operator in a fuzzy statistical database preserves fuzzy statistical dependency.

3.3. Fuzzy Statistical Projection Operator

Consider a fuzzy statistical table with scheme $FS(F_r, F_c, C)$. Let P1 be the path from root to leaf in F_r and P2 be the path from root to leaf in F_c then projection operator

$$Proj_{(P1,P2)}(FS)$$

will produce the fuzzy statistical table FS1 with P1 as fuzzy row attribute forest i.e. F_{r1} ,P2 as fuzzy column attribute forest i.e. F_{c1} whose fuzzy statistics C1 would be the fuzzy statistics corresponding to P1 and P2.

Example 4. Let FS12 be fuzzy statistical table

$FS12\big(state(sex(exp, salary)), (Incometax, Reward), (Count1, Count2)\big)$

Let P1 be the path $\theta_{r2} \to country \to sex \to exp$ and P2 be the path $\theta_{c2} \to Incometax$.

Then the operation

$$Proj_{(P1,P2)}(FS12)$$

produces the fuzzy statistical table FS

$FPS1\big(state(sex(exp)), (Incometax), (Count1)\big)$

where fuzzy statistics Count1 would be the fuzzy statistics corresponding to P1 and P2.

3.4. Fuzzy Statistical Table formation from several Fuzzy Primitive Statistical Tables

If either fuzzy row attribute tree or fuzzy column attribute tree of fuzzy primitive statistical tables are same then they can be combined to form a fuzzy statistical table. Let FPS_1 and FPS_2 be two fuzzy primitive statistical tables. If fuzzy column attribute tree of both are same then locate the node in fuzzy row attribute tree of FPS_1 which is different from the node in fuzzy row attribute tree of FPS_2 starting from the dummy node θ_{r1}. If node next to dummy node is different then we can use the concatenation operation. If not then locate the node which is different say B. Let A be the predecessor of B. Link the leaf node of FPS_2 to A next to B forming F_r of fuzzy statistical table FS.Depending upon the root to leaf path in F_r and F_c ,fuzzy statistics are taken care of. Similarly, if fuzzy row attribute tree are same then we can proceed for the column tree. In general, if $\{FPS_i : 1 \leq n\} = K$ is an ordered set of fuzzy primitive statistical tables. Then the operation

$$FS_{(F_r F_c)}(FPS_1, FPS_2, \ldots . FPS_n)$$

forms a fuzzy statistical table FS with F_r and F_c as its fuzzy row and column attribute forests and fuzzy statistics of $FPS_1, FPS_2, \ldots . FpS_n$ corresponding to the root to leaf path in F_r and F_c .

Example 5. Consider the fuzzy primitive statistical table $FPS1$ and $FPS2$ with scheme $FPS1\Big(\big(state(sex(exp))\big), (Incometax), (Count1)\Big)$
and $FPS2\Big(\big(state(sex(sal))\big), (Incometax), (Count2)\Big)$ .Then the operation

$$FS_{(F_r F_c)}(FPS1, FPS2)$$

creates a fuzzy statistical table $FS1\Big(\big(state(sex(exp, sal))\big), (Incometax), (count)\Big)$.

3.5. Decomposing a Fuzzy Statistical Table into Fuzzy Primitive Statistical Tables

For each X root to leaf path in F_r and for each Y root to leaf path in F_c and the corresponding fuzzy statistics in the fuzzy statistical table defines a fuzzy primitive statistical table. Let TRi (i=1…n) denotes the trees of F_r , TCj (j=1….m) denotes the trees of F_c , Ri be the root node of TRi, Cj be the root node of TCj. Then number of fuzzy primitive statistical tables in fuzzy statistical table FS is given by

$$\sum_{i=1}^{n} \mathrm{LN(Ri)} \times \sum_{j=1}^{m} \mathrm{LN(Cj)}$$

where LN(Ri) denote the number of leaf nodes of TRi and LN(Cj) denote the number of leaf nodes of TCj defined as:

$$\mathrm{LN(Ri)} = 1 \quad \text{if Ri is a leaf node}$$
$$= \sum_{i=1}^{n} \mathrm{LN(RSi)} \quad \text{otherwise}$$

where RSi are immediate successors of Ri. Similarly LN(Cj) is defined.

Let M denote the ordering among the fuzzy primitive statistical table of fuzzy statistical table FS by row-by-row enumeration of cells. Then, for fuzzy statistical table FS, composed of r fuzzy primitive statistical table FPS, the FPSi refers to the FPS at the i^{th} position in M. The operation $DEC_i(FS)$ returns the FPSi of FS.

Example 6. In fuzzy statistical table $FS1\Big(\big(state(sex(exp, sal))\big), (Incometax), (Count1)\Big)$ the operation $Dec_2(FS1)$ extracts the fuzzy primitive statistical table FPS2.

3.6. Extract Fuzzy Statistical Table

It is the inverse of concatenation. Suppose, in fuzzy statistical table FS, RT and CT be sets of integers denoting set of trees in F_r and F_c of FS and C be the ordered multiset of fuzzy statistics of FS. Then, the operation

$$EX_{(RT,CT)}(FS)$$

produces a fuzzy statistical table whose fuzzy row and fuzzy column attributes corresponds to the fuzzy attribute referenced in RT and CT. For example, consider the fuzzy statistical table

FS12
$FS12(state(sex(exp, salary)), (Incometax, Reward), (Count1, Count2))$
Here,
$RT = \{1\}$
$CT = \{1,2\}$
$C = \{Count1, Count2\}$
Then

$$EX_{[\{1\},\{1\}]}(FS12)$$

will produce the fuzzy statistical table $FS1(\text{state}(\text{sex}(\text{exp}, \text{salary})), (Incometax), (Count1))$.

3.7. Fuzzy Attribute Split in Fuzzy Statistical Table

This operation does not eliminate any cell values. It only relocates rows or columns of the fuzzy statistical table. Consider a fuzzy statistical table FS with fuzzy row attribute forest F_r and column attribute forest F_c. Let T with root A be a subtree of tree TR in F_r. Assume that A has k, k>1 immediate descendants and P denote the path of attributes from the root of TR to A. Then, the row split operation

$$SPILT_{(R,TR,P)}(FS)$$

maps T into k trees in which A is replaced by k new fuzzy attributes each named A and each having exactly one descendent of the splitted fuzzy attribute A as its child in the original order. The first subscript can only be row(R) or column(C) specifying TR $\in F_r$ or TC $\in F_c$ respectively. This operation does not eliminate any cell value, it only relocates rows of the fuzzy statistical table.

Example 7. Consider the fuzzy statistical table,
$FS1(\text{state}(\text{sex}(\text{exp}, \text{salary})), (Incometax), (Count1))$
the operation

$$SPLIT_{(R,1,country \rightarrow sex)}(FS1)$$

produces the fuzzy statistical table FSA shown in table 5.

3.8. Fuzzy Attribute Merge in Fuzzy Statistical Table

After splitting the fuzzy attributes, to regain the original fuzzy statistical table, the operation merge is used. Let B_1 and B_2 be the nodes with root A in trees T_1 and T_2 of F_r which are to be merged. Let P denote the path from root of T_1 to A which is also the path from the root of T_2 to A. The merge operation will merges B_2 to B_1 by making B_2 as a subtree of node A in tree T_1. The operation

$MERGE_{(R,T,P,B_1,B_2)}(FSA)$
where $T - \{T_1, T_2\}$
$MERGE_{(R,\{T_1,T_2\},(country \rightarrow sex),B_1,B_2)}(FSA)$
on fuzzy statistical table FSA produces a fuzzy statistical table FS. Similarly the column merge operation can be performed. When T and P are not specified then B_1 and B_2 are root nodes of distinct trees.

Example 7. Consider the fuzzy statistical table FSA after applying the spilt operation to FS1 in example 5. Then the operation

$$MERGE_{(R,\{T_1,T_2\},(country \rightarrow sex),B_1,B_2)}(FSA)$$

produces a fuzzy statistical table FS1.

4. IMPLEMENTING FUZZY STATISTICAL TABLE OPERATIONS

The objective of this section is to show briefly how the fuzzy statistical table operations are performed using the storage techniques described in [28].Here, we discuss concatenation, extract and merge operations. The other operations are similar.

4.1. Concatenation

To obtain the row concatenation of two fuzzy statistical table instances $FS1$ and $FS2$,in this order, we append the fuzzy statistics of $FS2$ to that of $FS1$. For the column concatenation of two fuzzy statistical table instances $FS1$ and $FS2$,the n^{th} row of the fuzzy statistics of FS2 is appended to the n^{th} row of that of FS1. The complexity of row concatenation and column concatenation is O(C(θ_r) $\times C(\theta_c)$).

4.2. Extract

Let FS be a fuzzy statistical table instance,T1 be a subtree of θ_r and T2 be the subtree of θ_c . Assume X and Y respectively be the roots of T1 and T2. The process of extracting the fuzzy statistical table associated with the row subtrees T1 and the column subtree T2 consists of locating in the original fuzzy statistics the row corresponding to the first root to leaf path instances of T1 ,reading only those fuzzy statistics corresponding to T1 and then writing them into the fuzzy statistics array which contains the fuzzy statistics of the fuzzy statistical table extracted. This process is repeated for each root to leaf path instance of T1. The complexity of this process is O(C(X) $\times$ C(Y)) .

4.3. Merge

To perform the column merge operation we read the rows of the original fuzzy statistics array from the first row to the last row, one row at a time. Once, a complete row has been read we relocate the cell attributes in the row according to the column tree produced by the column merge operation and write the resulting row into the fuzzy statistics of new fuzzy statistical table. Similarly, the row merge operation is performed by copying the rows of the old fuzzy statistic array into a fuzzy statistics array of new fuzzy statistical table, in the order established by the new row tree.

5. CONCLUSION

In this paper, we describe an approach for manipulation of fuzzy statistical tables. We define the fuzzy statistical join in fuzzy statistical framework and showed that fuzzy statistical dependency is preserved. Projection and certain basic operations are also defined. The performance of fuzzy statistical table operations is discussed. These findings offer important insights into the retrievability of information from a fuzzy statistical database.

ACKNOWLEDGEMENTS

We thank the School of Computer & Systems Sciences, Jawaharlal Nehru University for providing us the resources to conduct this research. We also thank all the researchers whose contribution to the field of fuzzy databases has helped us to give this paper the present shape.

Table 1. An instance of type-1 fuzzy statistical table 2000COUNT of highly salaried, highly paying incometax and highly experienced employees in a sample of a population

2000COUNT						Incometax				μ			
						2000	3000	4000	5000				
				Exp	4	15	79	87	34	0.25	0.33	0.4	0.4
					6	70	35	58	17	0.25	0.33	0.5	0.5
					8	90	98	84	35	0.25	0.33	0.5	0.66
			M		12	89	57	65	31	0.25	0.33	0.5	1
				Sal	50,000	54	56	30	35	0.25	0.33	0.5	0.67
					70,000	36	23	52	36	0.25	0.33	0.5	1
					80,000	87	90	92	94	0.25	0.33	0.5	1
	Delhi	Sex			90,000	67	20	72	54	0.25	0.33	0.5	1
				Exp	4	34	59	64	49	0.25	0.33	0.4	0.4
					6	67	18	44	75	0.25	0.33	0.5	0.5
					8	67	78	56	56	0.25	0.33	0.5	0.66
			F		12	54	94	49	94	0.25	0.33	0.5	1
				Sal	50,000	67	76	96	76	0.25	0.33	0.5	0.67
					70,000	98	29	62	56	0.25	0.33	0.5	1
					80,000	71	50	35	65	0.25	0.33	0.5	1
State					90,000	43	27	13	65	0.25	0.33	0.5	1
				Exp	4	86	60	93	57	0.25	0.33	0.4	0.4
					6	75	62	26	56	0.25	0.33	0.5	0.5
					8	67	57	12	65	0.25	0.33	0.5	0.66
			M		12	60	83	93	96	0.25	0.33	0.5	1
				Sal	50,000	68	29	46	57	0.25	0.33	0.5	0.67
					70,000	67	18	32	76	0.25	0.33	0.5	1
					80,000	34	73	47	56	0.25	0.33	0.5	1
	Bombay	Sex			90,000	23	70	13	66	0.25	0.33	0.5	1
				Exp	4	67	31	41	86	0.25	0.33	0.4	0.4
					6	56	27	87	76	0.25	0.33	0.5	0.5
					8	88	69	84	67	0.25	0.33	0.5	0.66
			F		12	43	10	63	87	0.25	0.33	0.5	1
				Sal	50,000	90	37	34	78	0.25	0.33	0.5	0.67
					70,000	84	92	62	56	0.25	0.33	0.5	1
					80,000	56	37	34	96	0.25	0.33	0.5	1
					90,000	58	20	96	65	0.25	0.33	0.5	1

Table 2 An instance of type-2 fuzzy statistical table FS1 in a sample of a population

FS1					Incometax				μ			
					3000	High	Low	4000-7000				
			Exp	10	5	70	60	10	1	0.59	0.63	1
				15-20	23	80	50	90	1	0.77	0.8	1
				Little	20	56	17	34	0.08	0.03	0.04	0.08
		M		Mod	21	45	67	56	0.5	0.33	0.25	0.2
			Sal	30,000	20	43	45	35	1	0.3	0.6	1
				High	10	56	78	56	0.47	0.67	0.41	0.5
				Low	45	56	57	68	0.68	0.59	0.8	0.2
Delhi	Sex			40,000-60,000	24	55	45	34	1	0.77	0.33	1
			Exp	10	32	25	63	46	1	0.2	0.63	1
				15-20	11	35	56	57	1	0.37	0.8	1
				Little	56	75	57	78	0.01	0.02	0.04	0.08
		F		Mod	20	56	63	34	0.14	0.17	0.2	0.09
			Sal	30,000	21	34	73	54	1	0.3	0.7	1

					High	11	64	47	56	0.34	0.29	0.32	0.36
					Low	23	54	24	45	0.64	0.36	0.6	0.5
State					40,000-60,000	45	34	64	57	1	0.26	0.23	1
				Exp	10	16	56	77	66	1	0.3	0.64	1
					15-20	46	56	25	78	1	0.67	0.7	1
					Little	59	86	63	87	0.01	0.08	0.04	0.03
			M		Mod	30	65	56	88	0.25	0.33	0.17	0.2
				Sal	30,000	19	56	36	56	1	0.4	0.41	1
					High	13	45	67	36	0.43	0.33	0.4	0.29
					Low	20	56	75	67	0.2	0.36	0.58	0.43
	Bombay	Sex			40,000-60,000	30	44	67	34	1	0.37	0.41	1
				Exp	10	25	66	78	46	1	0.5	0.7	1
					15-20	41	66	67	43	1	0.2	0.23	1
					Little	67	54	45	77	0.08	0.04	0.03	0.01
			F		Mod	46	67	53	67	0.14	0.25	0.5	0.17
				Sal	30,000	64	47	58	79	1	0.3	0.64	1
					High	20	12	84	42	0.29	0.3	0.4	0.47
					Low	45	85	68	34	0.5	0.62	0.67	0.53
					40,000-60,000	56	67	86	64	1	0.3	0.41	1

Table 3 An instance of fuzzy primitive table FPS1 in a sample of a population

						Incometax							
		FPS1				3000	High	Low	4000-7000		μ		
			M	Exp	10	5	70	60	10	1	0.59	0.63	1
					15-20	23	80	50	90	1	0.77	0.8	1
					Little	20	56	17	34	0.08	0.03	0.04	0.08
	Delhi	Sex			Mod	21	45	67	56	0.5	0.33	0.25	0.2
			F	Exp	10	32	25	63	46	1	0.2	0.63	1
					15-20	11	35	56	57	1	0.37	0.8	1
					Little	56	75	57	78	0.01	0.02	0.04	0.08
State					Mod	20	56	63	34	0.14	0.17	0.2	0.09
			M	Exp	10	16	56	77	66	1	0.3	0.64	1
					15-20	46	56	25	78	1	0.67	0.7	1
					Little	59	86	63	87	0.01	0.08	0.04	0.03
	Bombay	Sex			Mod	30	65	56	88	0.25	0.33	0.17	0.2
			F	Exp	10	25	66	78	46	1	0.5	0.7	1
					15-20	41	66	67	43	1	0.2	0.23	1
					Little	67	54	45	77	0.08	0.04	0.03	0.01
					Mod	46	67	53	67	0.14	0.25	0.5	0.17

Table 4 An instance of fuzzy primitive table FPS2 in a sample of a population

						Incometax							
		FPS2				3000	High	Low	4000-7000		μ		
			M	Sal	30,000	20	43	45	35	1	0.3	0.6	1
					High	10	56	78	56	0.47	0.67	0.41	0.5
					Low	45	56	57	68	0.68	0.59	0.8	0.2
	Delhi	Sex			40,000-60,000	24	55	45	34	1	0.77	0.33	1
			F	Sal	30,000	21	34	73	54	1	0.3	0.7	1
					High	11	64	47	56	0.34	0.29	0.32	0.36
					Low	23	54	24	45	0.64	0.36	0.6	0.5
State					40,000-60,000	45	34	64	57	1	0.26	0.23	1
			M	Sal	30,000	19	56	36	56	1	0.4	0.41	1
					High	13	45	67	36	0.43	0.33	0.4	0.29
					Low	20	56	75	67	0.2	0.36	0.58	0.43

	Bombay	Sex			40,000-60,000	30	44	67	34	1	0.37	0.41	1
			F	Sal	30,000	64	47	58	79	1	0.3	0.64	1
					High	20	12	84	42	0.29	0.3	0.4	0.47
					Low	45	85	68	34	0.5	0.62	0.67	0.53
					40,000-60,000	56	67	86	64	1	0.3	0.41	1

Table 5 An instance of fuzzy statistical table FSA in a sample of a population

FSA						Incometax							
						3000	High	Low	4000-7000	μ			
			M	Exp	10	5	70	60	10	1	0.59	0.63	1
					15-20	23	80	50	90	1	0.77	0.8	1
					Little	20	56	17	34	0.08	0.03	0.04	0.08
	Delhi	Sex			Mod	21	45	67	56	0.5	0.33	0.25	0.2
			F	Exp	10	32	25	63	46	1	0.2	0.63	1
					15-20	11	35	56	57	1	0.37	0.8	1
					Little	56	75	57	78	0.01	0.02	0.04	0.08
State					Mod	20	56	63	34	0.14	0.17	0.2	0.09
			M	Exp	10	16	56	77	66	1	0.3	0.64	1
					15-20	46	56	25	78	1	0.67	0.7	1
					Little	59	86	63	87	0.01	0.08	0.04	0.03
	Bombay	Sex			Mod	30	65	56	88	0.25	0.33	0.17	0.2
			F	Exp	10	25	66	78	46	1	0.5	0.7	1
					15-20	41	66	67	43	1	0.2	0.23	1
					Little	67	54	45	77	0.08	0.04	0.03	0.01
					Mod	46	67	53	67	0.14	0.25	0.5	0.17
			M	Sal	30,000	20	43	45	35	1	0.3	0.6	1
					High	10	56	78	56	0.47	0.67	0.41	0.5
					Low	45	56	57	68	0.68	0.59	0.8	0.2
	Delhi	Sex			40,000-60,000	24	55	45	34	1	0.77	0.33	1
			F	Sal	30,000	21	34	73	54	1	0.3	0.7	1
					High	11	64	47	56	0.34	0.29	0.32	0.36
					Low	23	54	24	45	0.64	0.36	0.6	0.5
State					40,000-60,000	45	34	64	57	1	0.26	0.23	1
			M	Sal	30,000	19	56	36	56	1	0.4	0.41	1
					High	13	45	67	36	0.43	0.33	0.4	0.29
					Low	20	56	75	67	0.2	0.36	0.58	0.43
	Bombay	Sex			40,000-60,000	30	44	67	34	1	0.37	0.41	1
			F	Sal	30,000	64	47	58	79	1	0.3	0.64	1
					High	20	12	84	42	0.29	0.3	0.4	0.47
					Low	45	85	68	34	0.5	0.62	0.67	0.53
					40,000-60,000	56	67	86	64	1	0.3	0.41	1

Table 6 An instance of fuzzy statistical table FSAGG in a sample of a population

						Incometax				μ					
	FSAGG					3000	High	Low	4000-7000						
				Exp	10	5	70	60	10	1	0.59	0.63	1	145	0.59
					15-20	23	80	50	90	1	0.77	0.8	1	243	0.77
					Little	20	56	17	34	0.08	0.03	0.04	0.08	127	0.03
			M		Mod	21	45	67	56	0.5	0.33	0.25	0.2	189	0.2
				Sal	30,000	20	43	45	35	1	0.3	0.6	1	143	0.3
					High	10	56	78	56	0.47	0.67	0.41	0.5	200	0.41
					Low	45	56	57	68	0.68	0.59	0.8	0.2	226	0.2
	Delhi	Sex			40,000-60,000	24	55	45	34	1	0.77	0.33	1	158	0.33
				Exp	10	32	25	63	46	1	0.2	0.63	1	166	0.2
					15-20	11	35	56	57	1	0.37	0.8	1	159	0.37
					Little	56	75	57	78	0.01	0.02	0.04	0.08	266	0.01
			F		Mod	20	56	63	34	0.14	0.17	0.2	0.09	173	0.09
				Sal	30,000	21	34	73	54	1	0.3	0.7	1	182	0.03
					High	11	64	47	56	0.34	0.29	0.32	0.36	178	0.29
					Low	23	54	24	45	0.64	0.36	0.6	0.5	146	0.5
State					40,000-60,000	45	34	64	57	1	0.26	0.23	1	200	0.23
				Exp	10	16	56	77	66	1	0.3	0.64	1	215	0.3
					15-20	46	56	25	78	1	0.67	0.7	1	205	0.67
					Little	59	86	63	87	0.01	0.08	0.04	0.03	295	0.01
			M		Mod	30	65	56	88	0.25	0.33	0.17	0.2	239	0.17
				Sal	30,000	19	56	36	56	1	0.4	0.41	1	167	0.4
					High	13	45	67	36	0.43	0.33	0.4	0.29	161	0.29
					Low	20	56	75	67	0.2	0.36	0.58	0.43	218	0.2
	Bombay	Sex			40,000-60,000	30	44	67	34	1	0.37	0.41	1	175	0.37
				Exp	10	25	66	78	46	1	0.5	0.7	1	215	0.5
					15-20	41	66	67	43	1	0.2	0.23	1	217	0.2
					Little	67	54	45	77	0.08	0.04	0.03	0.01	243	0.01
			F		Mod	46	67	53	67	0.14	0.25	0.5	0.17	233	0.14
				Sal	30,000	64	47	58	79	1	0.3	0.64	1	248	0.3
					High	20	12	84	42	0.29	0.3	0.4	0.47	158	0.29
					Low	45	85	68	34	0.5	0.62	0.67	0.53	232	0.5
					40,000-60,000	56	67	86	64	1	0.3	0.41	1	273	0.3

REFERENCES

[1] Zadeh, L.A. (1965) ,Fuzzy Sets,Inform. Control 8,338-353.

[2] Sato, H.(1981),"Handling Summary Information in a Database:Derivability", In Proceedings of the ACM SIGMOD International Conference on Management of Data.

[3] Buckles, B. P. & Petry, F. E. (1982), "A fuzzy representation for relational databases", Fuzzy Sets Syst. 7,213-226.

[4] Shoshani, A.,(1982)"Statistical databases: characteristics problems and some solutions", In Proceedings of the 8th International Conference on Very Large Data Bases ,pp. 208-222.

[5] Umano, M. (1982), Freedom-O, A fuzzy database system, In Fuzzy Information and Decision Processes, M.M. Gupta, E. Sanchez, Eds North Holland,Amsterdam, 337-347.

[6] Rafanelli, M. & Ricci, F.L.(1983) ,"Proposal of a model for statistical database",in Proc. Int Workshop Statistical Database Management, Los Altos, CA, Sept.27-29, pp 264-272.

[7] Chan, P. et al.(1983),Statistical data management research at Lawernce Berkley Laboratory,In Proc. 2nd Int.Workshop Statistical Database Management,Los Altos, CA, Sept. 27-29, pp 273-279.

[8] Ghosh, S. P.(1985),"An application of statistical database in manufacturing testing", IEEE Trans.Software Eng.,vol. SE 11,no.7, pp. 591-598, ; also IBM Res. Rep. RJ 4055, 1983.

[9] Shoshani, A. ,Olken, F. & Wong H.K.T.(1984),"Characteristics of scientific databases",Lawrence Berkley Lab Univ. California , Berkley ,Tech. Rep.LBL-17582.

[10] Prade , H. & Testemale, C.(1984)Generating database relational algebra for the treatment of incomplete or uncertain information and vague queries, Inf. Sci. 34,115-143.

[11] Kandel,A. & Zemankova-Leech, M.(1984),A fuzzy relational databases-A key to expert system,Verlag TUV, Rhineland Cologne.

[12] Ghosh, S.P.(1986)," Statistical Relational Tables for Statistical Database Management", IEEE Transactions on Software Engineering,vol. 12, no. 12,pp. 1106-1116.Also published as IBM Research Report RJ4394.

[13] Raju, K.V.S.V.N. & Majumdar,A.K.(1987) "The study of joins in fuzzy relational databases, Fuzzy sets and systems", vol 21, 19-34.
[14] Bhattacharjee T.K and Mazumdar A.K(1988),"Axiomatisation of fuzzy multivalued dependencies in a fuzzy relational data model", Fuzzy sets and systems , 343 – 352.
[15] Michalewicz, Z.(1991),Statistical and Scientific Databases,Z.Michalewicz (Ed.), Ellis Horwood,New York.
[16] Sato, H.(1991),Statistical Data Models: From a Statistical table to a Conceptual Approach,Chapter 7 in Statistical and Scientific Databases,Z.Michalewicz(Ed.), Ellis Horwood,New York.
[17] Kersten, R. Paul(1995),"The Fuzzy median and the Fuzzy Mad ,In Proceedings of ISUMA-NAFIPS.
[18] Berlin, Wu , Hung T. Nguyen(2006), Fundamentals of Statistics with Fuzzy Data, Springer-Verlag New York.
[19] Casillas, J. & Sanchez, L.(2006),"Knowledge extraction from fuzzy data for estimating consumer behavior models" IEEE International Conference on Fuzzy System,16-21.
[20] Antova, L., Jansen, T., Koch, C., Olteanu, D. (2008),"Fast and simple processing of uncertain data", In: Proc. of ICDE 2008, pp. 983–992.
[21] Benjelloun, O., Das Sarma, A., Halevy, A., Widom, J.(2006) ULDBs," Databases with uncertainty and lineage.", In: Proc. VLDB 2006, pp. 953–964.
[22] Dalvi, N., Suciu, D.: Management of probabilistic data: Foundations and challenges In: Proc. of PODS 2007, pp. 1–12 (2007)
[23] Das Sarma, A., Benjelloun, O., Halevy, A., Widom, J.(2006).,"Working models for uncertain data.", In: Proc. of 22nd Int. Conf. on Data Engineering, ICDE.
[24] Eiter, T., Lukasiewicz, T., Walter, M(2000),"Extension of the relational algebra to probabilistic complex values. In: Schewe, K.-D., Thalheim, B. (eds.) FoIKS 2000. LNCS, vol. 1762, pp. 94–115,Springer Heidelberg.
[25] Green, T.J., Tannen, V. (2006), "Models for incomplete and probabilistic information",IEEE Data Eng. Bull. 29, 17–24.
[26] Lakshmanan, L., Leone, N., Ross, R., Subrahmanian, V.S.(1997), Probview: A flexible probabilistic system. ACM Trans. Database Syst. 22(3), 419–469 .
[27] R´e, C., Dalvi, N., Suciu, D.(2006), Query evaluation on probabilistic databases.IEEE Data Eng. Bull. 29, 25–31.
[28] Guglani, S. ,Katti, C.P..(2012) 'A logical modeling tool to fuzzy statistical database and its physical organization', communicated to International Journal of Intelligent Information and Database System.
[29] Guglani, S. ,Katti, C.P. and Saxena, P.C.(2012),"Fuzzy statistical dependency and normalisation in fuzzy statistical database", Int. J. Intelligent Information and Database Systems, Vol. 6, No. 4