FUZZY STATISTICAL DATABASE AND ITS PHYSICAL ORGANIZATION

S.Guglani[1], C.P. Katti[2] and P.C. Saxena[3]

School of Computer & System Sciences,
JawaharLal Nehru University, Delhi, India
[1]seema_guglani@yahoo.co.in
[2]cpkatti@yahoo.com
[3]premchand_saxena@yahoo.com

ABSTRACT

Today most of the database systems in used are based on the precise data. However, in the real world application, it is often partially known or imprecise in nature giving rise to fuzzy relational database which contains imprecise micro data. Aggregation or statistical operation applied on micro data generates macro data called fuzzy statistic. In fuzzy statistics apart from fuzzy mean, fuzzy median primitive operation of counting of tuples in fuzzy relational database is an important aspect. In this context, we investigated the various approaches to cardinality of fuzzy sets which have been studied in the literature (e.g. sigmacount) and introduce the concept of fuzzy cardinality in fuzzy relational database. The concept of fuzzy statistical database is developed, which allows us to deal with the storage of vagueness or impreciseness associated with the statistical data. Fuzzy statistical database is a collection of fuzzy statistical tables. Depending upon the complexity of domain of fuzzy attributes in fuzzy statistical table, type-1 fuzzy statistical table and type-2 fuzzy statistical table are defined. We discuss the physical organization of fuzzy statistical table and algorithm for its implementation is also developed.

KEYWORDS

Fuzzy sets, Fuzzy attribute, Fuzzy relation, Fuzzy statistics, Statistical database, Fuzzy cardinality

1. INTRODUCTION

Most conventional databases in use today are based on the relational model. Each relation in the database represents a proposition and each record in a relation is a statement such that it evaluates to 'true' for that proposition. It could be argued, however, that this required precision actually gives an insufficient representation of the world. The model is grounded in binary black-and-white but much of reality actually exists in shades of grey. As such, the conventional relational database model has limited usefulness. One area that illustrates this limitation is in the everyday, subjective language generally used to describe people. For instance, a person might be described as being “tall, with a wide face and very dark brown eyes”. This description would be difficult to represent under the conventional relational model because it uses descriptive words that are inherently imprecise, and also because differing communities (i.e. groups with internal agreement on subjective meanings of these terms) may describe the same person differently. Many authors have made proposals to model and handle databases involving uncertain data giving rise to fuzzy relational database. The last two decades have witnessed a blossoming of research on this topic ([1],[7],[9],[15],[17],[20],[34-39]).Even though most of the literature about uncertain databases uses probability theory as the underlying uncertainty model, some approaches rather rest on possibility theory ([7],[21]). The initial idea consisting in applying possibility theory to this issue goes back to the early 80"s[15].Possibility theory has following advantages over probability theory.

- the qualitative nature of the model makes easier the elicitation of the degrees attached to candidate values.
- in probability theory , the fact that the sum of the degrees from a distribution must equal 1 makes it difficult to deal with incompletely known distributions.

Our aim is not to demonstrate that the possibility theory based framework is better than the probabilistic one at modeling uncertain, imprecise databases, but that it constitutes an interesting alternative in as much as it captures a different kind of uncertainty (of a qualitative nature). For example, a person who witnesses a car accident and is not sure about the model of the car involved. In such a case, it seems reasonable to model the vagueness by means of a possibility distribution .e.g. {1/Santro, 0.8/Verna.0.7/Honda} rather than with a probability distribution. Fuzzy relational database theory extends the relational model to allow for the representation of imprecise data and thus, it provides a more accurate representation of the world, that it models. Statistics calculated over such type of fuzzy relational database leads to imprecise statistic. Such statistical inferences are no longer confined to the boundaries of classical statistics. Some authors ([30-31]) have extended the concept of classical statistics in fuzzy environment leading to fuzzy statistics ([26],[31]).Cardinality of a fuzzy set is an important aspect. For example, a teacher is asked how many tall students he has in one of his classes. Some possible answers are (a) about 4 (b) 4 (c) 3.52. This is about counting the elements of a fuzzy set. It can, probably, be argued in different ways, but it is quite clear that (c) is not the best possible answer. The best possible answer is, perhaps (a) and the next answer is (b). Answers (b) and (c) are arrived at by using a nonfuzzy concept of cardinality[49] while (a) corresponds to a fuzzy cardinality concept([40],[42],[44]). Answer (c) is typically obtained by using the concept of sigma count (which was first defined by [41]). Fuzzy relation is nothing but a fuzzy subset of cartesian product of domain of its attributes which motivates us to define fuzzy cardinality of a fuzzy relation. Fuzzy statistics cannot be stored in the existing framework of statistical database ([8],[10],[13],[14],[16],[18],[19],[22], [23],[28]) as these are designed for the efficient storage and convenient retrieval of precise statistical data. This motivates us to develop fuzzy statistical database. The fuzzy relation[21] and fuzzy set theory proposed by [1] provide a requisite mathematical framework for dealing with such extended data values (precise and imprecise data values). Such type of fuzzy statistical database is quite useful when it is used as a decision, in areas such as management, decision making, health care, economic planning and census data evaluation. In the quest for providing proper logical model to fuzzy statistical database, we introduce fuzzy statistical tables (where partially as well as imprecise (fuzzy) known data values are permitted). Depending upon the type of fuzzy attributes [33] in fuzzy statistical table, type-1 and type-2 fuzzy statistical tables are defined. The statisticians may not need to use the entire fuzzy statistical tables during the preliminary stage of data analysis of certain operations. To enhance responsiveness the statistician may base their preliminary analysis on a part of fuzzy statistical table which motivates us to define fuzzy primitive statistical table. The physical organization of data has a major impact on any database system performance because it is the level at which actual implementation takes place in physical storage. Fuzzy statistical table is a logical model for fuzzy statistical database, so in view of its storage in the system we propose its physical organization technique and algorithm for its implementation is developed.

The paper is organized as follows. In section 2, we introduce the preliminaries which include fuzzy sets and fuzzy relations. In section 3, fuzzy count is defined. In section 4, fuzzy statistical database is proposed and its logical model, fuzzy statistical table is introduced. In section 5, fuzzy primitive statistical tables are defined. In section 6, physical organization of fuzzy statistical table is discussed and finally in section 7, algorithm for implementation of fuzzy statistical table is developed.

2. PRELIMINARIS

2.1. Fuzzy Set

Fuzzy sets have been introduced by [1] as an extension of the classical notion of set. In classical set theory, the membership of elements in a set is assessed in binary terms according to a bivalent condition –an element either belongs or does not belong to the set. By contrast, fuzzy set theory ([26] ,[27],[29])permits the gradual assessment of the membership of elements in a set, which is described with the aid of a membership function valued in the real unit interval [0,1]. Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special cases of membership functions of fuzzy sets if the latter only take values 0 or 1.Let X be a classical set of objects called universe of discourse whose elements are denoted by x. A fuzzy set A in X is characterized by membership function μ_A:X→[0,1] where $\mu_A(x)$ for each x∈X denotes the grade of membership of x in the fuzzy set A. Such type of fuzzy sets are known as type-1. Following the notations used in fuzzy set theory, we write

$$A= \{\mu_A(x1) \,/\, x1 ,\ldots\ldots,\mu_A(x_n) \,/\, x_n\}$$

where $x_i \in X, 1 \leq i \leq n$. A membership function to the value 0 describes that the member is not included in the fuzzy set, 1 describes a fully included member of the fuzzy set. Value strictly between 0 and 1 characterize the fuzzy members, which belong to the fuzzy set partially. Depending upon the complexity of underlying domain, fuzzy sets may be of type-2.Type-2 fuzzy sets are defined to be fuzzy sets whose grades of membership are themselves fuzzy.

2.2 Fuzzy Relations

Fuzzy relations were developed to capture various types of imprecise information occuring in the real world. The imprecise values in a database system can be broadly classified into two types-

- Attribute Uncertainity : The situation where some attribute values are uncertain.
- Existential Uncertainity : The situation where the existence of some tuples is uncertain.

Fuzzy relation can very well incorporate both of these impreciseness. Out of the various approaches proposed in the fuzzy database literature to represent impreciseness in attribute values, we now discuss some of the significant ones. Broadly two approaches have been most popular. First is the similarity based approach, which characterizes the impreciseness by using linguistic terms e.g. poor, fair, good etc. and the degree of similarity between a pair of linguistic terms is characterized by a similarity matrix.The second approach is based on Zadeh‟s possibility theory[2]which uses possibility distribution as a value of an attribute to capture the impreciseness of the first type. The second type of impreciseness i.e. the partial membership of a tuple in a relation allows a tuple to be a partial member of a relation, for e.g. the animals which are considered "somewhat endangered" are partial members of the Endangered-Species relation. A tuple with a partial membership in a relation is referred as weighted tuple ([20],[21]). The possibility based approach [9] described above is believed to be more general and popular; the important reason being that it handles all types of imprecise information.

We now provide the basic notations and concepts in two possibility based fuzzy relational data models which we treat in this paper. The models are known as type-1 & type-2 fuzzy relational data models

[20].While introducing them, we will show how the models capture all types of impreciseness in a better way. Adhering to the notations of classical relational database theory, a relation schema R in fuzzy relational data model is defined as a finite set of attributes $\{A_1,A_2,\ldots A_n\}$ and is denoted as R $(A_1,A_2,\ldots A_n)$. Corresponding to each attribute A_i,i=1…n , is a set $dom(A_i)$ called the domain of A_i.Along with each attribute A_i, a set U_i, called universe of discourse U_i is associated for the domain values of A_i. A fuzzy relation r on a relation scheme R $(A_1,A_2,\ldots A_n)$ is a fuzzy subset of $dom(A_1)\times dom(A_2)\times\ldots\times dom(An)$. Mathematically, fuzzy relation r is defined as a fuzzy subset of $U_1\times U_2\times\ldots U_n$ and is characterized by the n-variate membership function μ_r; $U_1\times U_2\times\ldots U_n\rightarrow[0,1]$. Like classical relation, fuzzy relation r will be represented as a table with an additional column for $\mu_r(t)$ denoting the membership value of the tuple t in r. Depending upon the complexity of domain of attributes in fuzzy relation it is classified into two types : type-1 fuzzy relation and type-2 fuzzy relation.

In type-1 fuzzy relations, dom(A_i) is a fuzzy set(or a classical set). A type-1 fuzzy relation may be considered as a first-level extension of classical relations,where impreciseness in the association among entities can be captured.The type-2 fuzzy relations provide further generlization by allowing dom(A_i) to be a set of fuzzy sets.Table 1 shows an instance of type-2 fuzzy relation EMPLOYEE(Name,State,Sex,Experience,Salary,Incometax) in an organization which contains the information about name, state where the employee lives, sex(male or female), his expérience, his status of salary and the incometax paid by the employee. In the EMPLOYEE relation dom(Name), dom(State) and dom(Sex) are assumed to be crisp set while dom(Experience), dom(Salary) and dom(Incometax) are sets of fuzzy sets in their universes $U_{Experience}$, U_{Salary} and $U_{Incometax}$ respectively. $U_{Experience}$, U_{Salary} and $U_{Incometax}$ are assumed to be sets of positive integers in the range 0-30,10,000-100,000 and 0-10,000 respectively. The fuzzy set descriptors High, Low, Little and Moderate have been used to represent fuzzy data values over respective domains. The membership functions of the fuzzy set descriptors High, Low, Little and Moderate are domain dependent and are as given below :

For $x\in U_{Experience}$,

$$\mu_{Moderate}(x) = (1+|x-8|)^{-1} \text{ for } x > 1$$
$$= 0 \text{ otherwise}$$

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

Similarly,

$$\mu_{\mathrm{High}}(y) = (1 + a|y-c|)^{-1} \text{ for } y \le c$$
$$= 1 \text{ for } y > c$$

where a = 1/20,000 , c = 60,000 for $y \in U_{Salary}$ and a = 1/1,000 , c = 5,000 for $y \in U_{Incometax}$

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

3. FUZZY COUNT

As is often the case when extending a classical set theoretical notion to fuzzy set theory, various alternative definitions for the notion of the cardinality of a fuzzy set (on a finite universe) have been proposed. These proposals can be roughly subdivided in two categories:

i. scalar cardinalities, where the cardinality of a fuzzy set is a positive real number. ([34],[41],[45-47]).
ii. fuzzy cardinalities, where the cardinality of a fuzzy set is defined as a fuzzy quantity (not necessarily convex) ([1],[40],[47-49]).

Scalar cardinality of a fuzzy set is the sum of the membership values of all elements of the fuzzy set. In particular, scalar cardinalities of a fuzzy set associate to each fuzzy set a positive real number. In the similar way, fuzzy cardinalities of a fuzzy set associate to any fuzzy set a convex fuzzy natural number. The fuzzy cardinality of fuzzy sets is itself also a fuzzy set on the universe of natural numbers. The first definition of fuzzy cardinality of fuzzy sets by means of mapping from the set of natural numbers to the interval [0,1], was proposed by Zadeh ([3-4].As fuzzy relation r on a relation scheme $R\ (A_1,A_2,\ldots.A_n)$ is a fuzzy subset of $dom\ (A_1) \times dom\ (A_2) \times \ldots\ldots \times dom\ (A_n)$ which motivates us to define the fuzzy cardinality of a fuzzy relation. Let t_i be the i^{th} tuple of fuzzy relation r and $\mu_r(t_i)$ denote the membership degree of i^{th} tuple in r, then fuzzy cardinality of fuzzy relation r , called as fuzzy count (denoted by FC) is defined by

$$\mathrm{FC} = \begin{cases} 0 \ \ if\ r = \emptyset \\ j \ \ if \ \ r \neq \emptyset\ and\ \mu_r(t_j) \ \geq 0.5 \\ j-1 \ \ if \ \ r \neq \emptyset\ and\ \ \mu_r(t_j) \ < 0.5, \end{cases}$$

$$\text{where} \quad j = \max \left\{ \begin{matrix} 1 \leq s \leq n : \ \mu_r(t_{s-1}) + \mu_r(t_s) > 1 \ \ if\ r \neq \emptyset \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad if\ r = \emptyset \end{matrix} \right\}.$$

with membership degree denoted by μ_{Count} and defined as

$$\mu_{Count}=\mu_r\ (t_{FC}) \wedge (1-\mu_r\ (t_{FC}+1)\ )$$

using the representation of r as

$$r = \begin{pmatrix} t_1 & t_2 & \ldots & \ldots\ldots t_n \\ \mu_r(t_1) & \mu_r(t_2) & \ldots\ldots & .\ \mu_r(t_n) \end{pmatrix}$$

where $\mu_r\ (t_1)\ ,\ldots..\mu_r\ (t_n)$ arranged in decreasing order of magnitude and $\mu_r(\ t_0) = 1$, $\mu_r(\ t_{n+1}) = 0$.

The following simple procedure can be used to calculate fuzzy count of fuzzy relation r.

Algorithm

i. Input : a fuzzy relation r.
ii. Take fuzzy projection([3],[4]) and fuzzy selection ([3],[4]) on r over the specified predicates yielding fuzzy relation sr.
iii. If sr=$\emptyset$, set FC = 0 with $\mu_{Count}= \mu_r(t_0) \wedge (1-\mu_r(t_1)\ )$. Stop.
iv. Arrange the tuples of sr such that $\mu_r(t_1)\ ,\ldots..\mu_r(t_n)$ are arranged in decreasing order of magnitude.
v. Set pointer p to the first tuple of sr.
vi. If p $\neq nil$
   initialize A and B array
   i←1
   while (p $\neq nil$)
      A[i]←$\mu_{sr}(t_i-1)+\mu_{sr}(t_i)$
      i++
      p++

k←1
largest←0
for j←1 to (i–1)
if (A[j]>1) then
B[k]←j
k++
Store the largest element of B in largest

vii. If $\mu_{sr}(t_{largest}) \geq 0.5$ then FC = largest.
viii. If $\mu_{sr}(t_{largest}) < 0.5$ then FC = largest − 1.
ix. $\mu_{count}=\mu_{sr}\ (t_{FC}) \wedge (1-\mu_{sr}\ (t_{FC+1})\ )$.
x. Stop.

Example. Consider a fuzzy relation EMPLOYEE shown in table 1 as discussed in section 2.2. If the query is: find the number of male people in Delhi who are Little experienced and paying low incometax then the answer to it is obtained by using the algorithm of fuzzy count in EMPLOYEE fuzzy relation as follows:

1. Take fuzzy projection on EMPLOYEE over the attributes State, Sex, Exp and Incometax.
2. Select the tuples(fuzzy selection) in the fuzzy relation obtained in step 1 where State=Delhi Sex=M, Exp=little and Incometax=low.
3. Represent sr as

$$sr = \begin{pmatrix} t_1 & t_2 & t_3 & t_4 \\ 0.7 & 0.7 & 0.5 & 0.5 \end{pmatrix}$$

4. Using step (v)-(viii) of above algorithm ,we get FC=3
5. Using step (ix), we get μ_{Count}=0.05

The answer to the query is 3 with membership degree 0.05. We can also say that the number of male people in Delhi who are little experienced and paying low incometax is 3 with truth value 0.05.
Our définition of fuzzy count is in accordance with that of classical cardinality

Preposition 1. If r is a classical relation with k tuples then fuzzy count of r is k.
Proof. If r is a classical relation with k tuples then r can be represented as

$$r = \begin{pmatrix} t_1 & t_2 & \ldots & \ldots\ldots t_k \\ \mu_r(t_1) & \mu_r(t_2) & \ldots\ldots & \ldots\mu_r(t_k) \end{pmatrix}$$

where $\mu_r(t_1)$,…..$\mu_r\ (t_k)$ arranged in decreasing order of magnitude and $\mu_r(t_0)$ =1,$\mu_r\ (t_{k+1})$ =0. Using the above algorithm,

$$FC=k$$
$$\mu_{Count} = \mu_{sr}\ (t_k) \wedge (1-\mu_{sr}\ (t_{k+1})).$$
$$=1 \wedge (1-0) = 1.$$

Hence fuzzy count of r is 1 with truth value 1.

Proposition 2. Fuzzy cardinality of a fuzzy relation is a fuzzy convex set.
Proof. Let r be a fuzzy relation.Then,

$$\mu_r(t_k) \wedge \left(1-\mu_r(t_{k+1})\right) = \begin{cases} \mu_r(t_k) & if \;\; \mu_r(t_k)+\mu_r(t_{k+1}) \le 1 \\ 1-\mu_r(t_{k+1}) & \text{otherwise} \end{cases}$$

Then there is a unique value j such that following inequalities hold :

$$1+\mu_r(t_1) \ge \mu_r(t_1)+\mu_r(t_2) \ge \mu_r(t_2)+\mu_r(t_3) \ge \cdots\cdots \ge \mu_r(t_{j-1})+\mu_r(t_j) > 1 \ge \mu_r(t_j)+\mu_r(t_{j+1}) \ge \cdots\cdots \ge \mu_r(t_{n-1})+\mu_r(t_n) \ge \mu_r(t_n) \ge 0$$

Let FC = k, then

$$\mu_{count} = \begin{cases} 1-\mu_r(t_{k+1}) & for\; 0 \le k \le j-1 \\ \mu_r(t_k) & for\; j \le k \le n \end{cases}$$

i.e. $\quad 1-\mu_r(t_1) \le \cdots \le 1-\mu_r(t_j) \quad and \quad \mu_r(t_j) \ge \mu_r(t_{j+1}) \ge \cdots \ge \mu_r(t_n)$

$$\Rightarrow max(1-\mu_r(t_{j+1}), \mu_r(t_j)) = \begin{cases} \mu_r(t_j) & if \;\; \mu_r(t_j) \ge 0.5 \\ 1-\mu_r(t_{j+1}) & \text{otherwise} \end{cases}$$

The pattern of membership degrees of being increasing up to a certain value i.e. $1-\mu_r(t_{j+1})$ or $\mu_r(t_j)$ and then decreasing signifies that fuzzy cardinality of a fuzzy relation is a fuzzy convex set.

4. FUZZY STATISTICAL DATABASE

A statistical database is a database used for statistical analysis purposes. Over the last several years there has been growing interest in statistical database. This interest is due to the inadequacy of commercial database management systems to support statistical applications. A similar situation exists in other application areas such as CAD/CAM, VLSI design, and knowledge-based systems. The main reason for this situation is that most of the commercial systems available today were designed primarily to support transactions for business applications (such as marketing and banking), while other applications has different data characteristics and processing requirements. Another reason for the interest is the large amount of data that exist in statistical database applications. Many practical databases collected for statistical purposes, such as trade data between countries, or the various census data, are too large to be handled with conventional data management techniques efficiently. But, in addition, there is a large amount of data that was not collected originally for experimental or statistical purposes, that has tremendous potential when used for statistical purposes. For example, routine patient records in hospitals can be used for statistical "cause and effect" studies. Business transactions can be statistically analyzed for policy setting and econometric models. For the most part, such sources of routine collections of data are left unused because adequate data management and analysis facilities do not exist. Statistical database applications differ from commercial applications both in the properties of the data and in the operations over the data. For example, statistical database often contain sparse data (usually in a multidimensional matrix form) that need to be efficiently compressed. They often contain special' data types such as vectors or time series. Similarly, statistical analysis requires different types of operations over the data. Typically, only a few variables are examined over a large number of the cases (as is the case in determining cross-correlation). In addition to statistical operators, such as sampling and aggregation, the access of the data is of a different nature. For example, it is quite common to access a region in multidimensional space, such as finding materials with certain approximate properties, or cases that fit a statistical pattern. For such cases multidimensional data structures and search methods are desirable. A statistical database can be thought of as a collection of data sets which gives summary information on a certain population of individuals or objects. It contains macro data, a collection of micro data, obtained by statistical inferences over relational databases ([12],[19],[32]). Many researchers have explored the fundamentals of statistical database([6],[8],[10-14],[16],[18],[22-24],[28]) could not be stored in the existing framework of statistical database as these are designed for storing the precise statistical data. It is therefore imperative to provide a data model[5] for the storage of fuzzy statistics which motivates us to develop the concept of fuzzy statistical database. A fuzzy statistical database gives the user the ability to obtain fuzzy statistics information directly from the fuzzy statistical database which relieves the user from the task of calculating fuzzy statistics from the fuzzy relation. 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.

A fuzzy statistical scheme in fuzzy statistical data model is defined as $FS(Fr,Fc,C)$ where

- Fr denote the fuzzy row attribute forest consisting of finite set of fuzzy attributes[33].
- Fc denote the fuzzy column attribute forest consisting of finite set of fuzzy attributes and
- C is the fuzzy statistics.

Fuzzy statistical table on fuzzy statistical scheme $FS(Fr,Fc,C)$ is a tabular representation of fuzzy statistics. It is represented as a table with additional two-dimensional array of cells for representing the membership degree of fuzzy statistics, denoted by μ .The row header and column header of a fuzzy statistical table are labeled by fuzzy attributes. Fuzzy attributes are structured in the form of an ordered set of trees i.e. Fr and Fc .Each cell in a fuzzy statistical table has an associated set of fuzzy row and 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. We now turn to a more formal explanation of our definition.

Fuzzy statistical database is a collection of fuzzy statistical tables. A parenthesized expression to specify an attribute tree which is a preorder enumeration of the tree (i.e. first the root then the sub trees from left to right) is used. Suppose for accessing the fuzzy statistic C in fuzzy statistical table, $RA_1,\ldots,RA_n$ be fuzzy row attributes with their appropriate universes $RU_1,\ldots,RU_n$ respectively which forms a path from root to a leaf in fuzzy row attribute tree of F_r and $CB_1,\ldots CB_m$ be fuzzy column attributes with their appropriate universes $CU_1,\ldots,CU_m$ which forms a path from root to a leaf in a fuzzy column attribute tree of Fc . Then fuzzy statistical table on fuzzy statistical scheme FS($RA_1(RA_2(RA_3\ldots\ldots RA_n)),(CB_1(CB_2\ CB_3\ldots CB_m$)),(C)) is a fuzzy subset of $RU_1\times RU_2\times\ldots\times RU_n\times CU_1\times CU_2\times\ldots\times CU_m$.

A fuzzy statistical table instance is a collection of cell instances structured as specified by the fuzzy statistical scheme. A cell instance consists of value of its fuzzy row and fuzzy column attributes 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: A type-1 fuzzy statistical table enables us to capture impreciseness of the statistics associated with the fuzzy row and fuzzy column attributes. In type-1 fuzzy statistical table the domain of fuzzy row and fuzzy column attributes may be a classical subset or a fuzzy subset of their appropriate universes. In other words, here fuzzy attributes [33] are of type-1.Let c be an instance of fuzzy statistics in two dimensional array along with its membership degree $\mu(c)$ and $RA_1,\ldots,RAn$ be fuzzy row attributes with their appropriate universes $RU_1,\ldots,RUn$ respectively which forms a path from root to a leaf in fuzzy row attribute tree for accessing c alongwith their membership function denoted by μ_{RAi},

$$\mu_{RAi}:RU_i \rightarrow [0,1] \; ,i = 1....n.$$

Let $CB_1,....CB_m$ 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 c along with their membership function denoted by μ_{CBj} for j=1….m,

$$\mu_{CBj}:CU_j \rightarrow [0,1]$$

Consider the Cartesian product, $dom\ (RA_1) \times dom\ (RA_2) \times....\times dom(RAn) \times dom(CB_1) \times dom\ (CB_2) \times.........\times dom(CBm)$.

It is a fuzzy subset of $RU_1 \times RU_2 \times......\times RUn \times CU_1 \times CU_2 \times.....\times CUm$. In type-1 fuzzy statistical table the membership function of fuzzy statistic satisfies the following inequality:

$$\mu\ c\ \leq \min(\ \mu_{RA1}\ (ru1)\ ,....\mu_{RAn}\ (ru_n)\ ,\mu_{CB1}\ (cu_1)\ ,..\mu_{CBm}\ (cu_m))$$

where $ru_i \in RU_i$, $cu_j \in CUj$, i=1…n, j =1….m. We can also treat $\mu(c)$ as a fuzzy truth value belonging to [0,1] i.e. $\mu(c)$ is a truth value of a fuzzy predicate associated with c when the fuzzy attributes which are occurring in path from root to leaf to access c are replaced by their values. We illustrate type-1 fuzzy statistical table by an example.

Example. Consider a fuzzy statistical scheme

2000COUNT(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 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 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. Table 2 shows an instance of fuzzy statistics table 2000COUNT. In 2000COUNT 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 $x \in U_{Exp}$,

$$\mu_{HX}(x) = (1 + |x - 10|/4)^{-1} \quad \text{for } x \leq 10$$
$$= 1 \quad \text{for } x > 10$$

For $s \in U_{Sal}$,

$$\mu_{HS}(s) = (1 + |s - 60{,}000|/20{,}000)^{-1} \quad \text{for } s \leq 60{,}000$$
$$= 1 \quad \text{for } s > 60{,}000$$

For $y \in U_{Incometax}$,

$$\mu_{HI}(y) = (1 + |y - 5{,}000|/1{,}000)^{-1} \quad \text{for } y \leq 5{,}000$$
$$= 1 \quad \text{for } y > 5{,}000$$

Type-2 Fuzzy Statistical Table: Although type-1 fuzzy statistical table enable us to represent impreciseness of the statistics associated with fuzzy attributes, its role in capturing uncertainty in data values is rather limited. For example, in a type-1 fuzzy statistical table 2000COUNT, one is not permitted to specify the salary of male people in Delhi to be in the range Rs. 40,000-Rs.50,000 or low because of type -1 fuzzy attributes. So in order to accommodate a wider class of data ambiguities, a further generalization of the type-1 fuzzy statistical table is considered in the form of type-2 fuzzy statistical table. In type-2 fuzzy statistical table domain of fuzzy row and fuzzy column attributes are sets of fuzzy sets in their appropriate universes. In other words it consists of type-2 fuzzy attribute. Suppose c be an instance of fuzzy statistics in type-2 fuzzy statistical table with membership degree μc and RA_1,…. RA_n be fuzzy row attributes with their appropriate universes RU_1,….,RU_n which forms a path from root to a leaf in a fuzzy row attribute tree for accessing c along with their membership function denoted by μ_{RAi} for i=1…n, μ_{RAi}:RUi→ 0,1 and $CB1$,.CBm be fuzzy column attributes with their appropriate universes $CU1$,…..,CUm respectively which forms a path from root to a leaf in a fuzzy column attribute tree for accessing c along with their membership function denoted by μCBj for j=1…m , μCBj:CUj→ 0,1 . Here, dom(RAi) is the set of fuzzy sets in RUi, i=1….n and dom(CBj) is the set of fuzzy sets in CUj, j=1….m. Consider a tuple t=(ra_1,ra_2,….,ra_n,cb_1,cb_2….,cbm) in cartesian product $dom\ RA_1 \times dom\ RA_2 \times .. \times dom\ RAn \times dom\ CB_1 \times dom\ CB_2 \times ...... dom(CBm)$. It is a fuzzy subset of $RU_1 \times RU_2 \times ... RUn \times CU_1 \times CU_2 \times .... CUm$. In type-2 fuzzy statistical table, the membership function of fuzzy statistic satisfies the following inequality:

$$\mu(c) \leq \max_{\substack{(ru_1,..ru_n,cu_1,..cu_m) \in \\ RU_1 \times .. \times RU_n \times CU_1 \times .. CU_m}} min(\mu_{ra_1}(ru_1),..,\mu_{ra_n}(ru_n),\mu_{cb_1}(cu_1),..,\mu_{cb_m}(cu_m))$$

$\mu(c)$ may be regarded as a fuzzy truth value belonging to [0,1] i.e. μ (c) is a truth value of a fuzzy predicate associated with c when the fuzzy attributes which are occurring in path from root to leaf to access c are replaced by their values. We illustrate type-2 fuzzy statistical table by an example.

Example . Consider a type-2 fuzzy statistical scheme

2001COUNT(State(Sex(Exp,Sal)),(Incometax),(Count))

in a sample of a population shown in table 3. As in above example , 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.

$dom\ (Exp)$ = set of fuzzy sets in U_{Exp} ={Little,Mod,10,15-20}
$dom\ (Sal)$ =set of fuzzy sets in U_{Sal} ={30,000,High,Low,40,000–60,000}
$dom\ (Incometax)$ =set of fuzzy sets in $U_{Incometax}$ ={High,Low,3,000,4,000–7,000}

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

For $x \in U_{Exp}$,

$$\mu_{Mod}(x) = (1 + |x - 8|)^{-1} \quad \text{for } x > 1$$
$$= 0 \quad \text{otherwise}$$
$$\mu_{Little}(x) = (1 + 12x)^{-1} \quad \text{for } x > 0$$
$$= 0 \quad \text{otherwise}$$

For $y \in U_{Sal}$,

$$\mu_{High}(y) = (1 + |y - 60{,}000|/20{,}000)^{-1} \quad \text{for } y \leq 60{,}000$$
$$= 1 \quad \text{for } y > 60{,}000$$
$$\mu_{Low}(y) = 1 - \mu_{High}(y)$$

For $y \in U_{Incometax}$,

$$\mu_{High}(y) = (1 + |y - 5{,}000|/1{,}000)^{-1} \quad \text{for } y \leq 5{,}000$$
$$= 1 \quad \text{for } y > 5{,}000$$
$$\mu_{Low}(y) = 1 - \mu_{High}(y)$$

The fuzzy statistics count 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.

5. FUZZY PRIMITIVE STATISTICAL TABLE

Fuzzy statistical table provides an efficient logical modelling tool for statisticians to store fuzzy statistics in a statistical environment. Sometimes statisticians are not interested in the entire fuzzy statistical table. Instead of taking full tables into account, their analysis may based on a part of fuzzy statistical table. For example, for making decisions on how many people to be retained in an organization the statistics of experienced people is to be taken care of. Then there is no need to go through the entire 2001COUNT fuzzy statistical table, a part of it serves the desired purpose. This motivates us to propose fuzzy primitive statistical table.

Definition. 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 III 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 are shown in table 4 and table 5 respectively.

6. PHYSICAL ORGANIZATION OF FUZZY STATISTICAL TABLE

If for any sub tree in fuzzy row attribute forest F_r or in fuzzy column attribute forest F_c , there is always a unique instance in a fuzzy statistical table, we say that the fuzzy statistical table has a full cross product otherwise it is incomplete. For example, consider the fuzzy statistical table FSINC instance in table 6. It does not have a full cross product as the instance of Exp as a sub tree of Delhi is different from the instance of Exp as a sub tree of Bombay. Assuming that the fuzzy statistical table have a full cross product, we separately store the fuzzy row attribute forest F_r, fuzzy column attribute forest F_c ,fuzzy statistics and their membership degree. The cell value arrays C_s and C_μ store the fuzzy statistics and their membership degrees respectively. The cell value array C_μ is in one to one correspondence with C_s.

6.1. Computing Fuzzy Attributes

For simplicity, we put fuzzy row forest F_r 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 into ordered tree TC by making the root nodes in F_c as an immediate descendents of dummy attribute θc.Consider an ordered sub tree of tree TR or TC with root R. The total number of leaves in the instance of the sub tree with root R denoted by NL(R) is given by

$$\begin{aligned} NL(R) &= |R| && \text{if R is a leaf} \\ &= |R| \textstyle\sum_{i=1}^{n} NL(S_i) && \text{otherwise} \end{aligned}$$

where $|R|$ denotes the number of instance of attributes R and S_1,S_2,……S_n are immediate successors of R

Let attribute S be an immediate successors of attribute R in an ordered tree. The total number of leaves in the instance of a sub tree with root R and to the left of S, denoted by $NL(R,S)$ is given by

$$\begin{aligned} NL(R,S) &= 0 && \text{if S is the leftmost immediate successor of R} \\ &= \textstyle\sum_{i=1}^{n} NL(S_i) && \text{otherwise} \end{aligned}$$

where S_1,S_2,……S_n are immediate successors of R and they are to the left of S in an ordered tree. Let $TNL(R)$ denote the total number of leaves in an instance corresponding to one sub tree of R, then

$TNL(R)$ =$NL(R)$ / $|R|$

For efficient representation of fuzzy attributes, we use the total ordering to the values of a fuzzy attribute and use the relative position of an attribute value as its encoded value. For example, consider Sex attribute and it's ordered instances M and F, we would use integers 0 and 1 to encode M and F respectively. Attributes with their instance and coded value are given in table 7.

Let (P,V) denote the root to leaf path instance such that

1. P = $(X1,X2,\ldots..Xn)$ is a root to leaf path of TR or TC where the root is omitted. X_1 is an immediate successor of the root and X_{i+1} is an immediate successor of X_i for $1 \leq i < n$
2. V= $(x1,x2,\ldots\ldots xn)$ is an ordered set of values where $x_i \geq 0$ is an encoded value for the attribute X_i , $1 \leq i \leq n$.

For TR, when root to leaf path instance is given (P,V) the corresponding row leaf number is denoted by RLN((X_1,x_1) , (X_2,x_2) ,….. (X_n,x_n) or simply RLN. It can be computed as

$$RLN\big((X_1,x_1),(X_2,x_2),\ldots..(X_n,x_n)\big) = \textstyle\sum_{i=1}^{n}(NL(X_{i-1},X_i) + x_i * TNL(X_i))$$

where X_0 is the dummy root attribute of the ordered tree. Leaves are numbered beginning at 0. Similarly, for TC the column leaf number denoted by CLN corresponding to root to leaf instance (P,V) can be computed from above formula.

Given N (row leaf number or column leaf number), the corresponding root to leaf path instance can be obtained as follows:

Let {S_1,S_2,……S_n} be the ordered set of immediate successors of X_0 (the dummy root attribute). Then

- X_1 is equal to S_k where
  NL(θ, S_k)≤N<NL(θ, S_{k+1}) if k≠n
  NL(θ, S_k)≤N if k=n
- The value for x_1 is given by
  $x_1 = \lfloor (N - NL(\theta, X_1))/TNL(X_1) \rfloor$

Now let $\{S_1, S_2, \ldots \ldots S_n\}$ be the ordered set of immediate successors of X_{j-1}, $j > 1$. Then

- X_j is equal to S_k where
  NL(X_{j-1}, S_k) ≤ N− $\sum_{i=1}^{j-1} NL(X_{i-1}, X_i) - \sum_{i=1}^{j-1} x_i\, TNL(X_i)$ < NL(X_{j-1}, S_{k+1}) if k≠n
  NL(X_{j-1}, S_k) ≤ N− $\sum_{i=1}^{j-1} NL(X_{i-1}, X_i) - \sum_{i=1}^{j-1} x_i\, TNL(X_i)$ if k=n
- The value for x_j is given by
  $x_j = \lfloor (N - \sum_{i=1}^{j} NL(X_{i-1}, X_i) - \sum_{i=1}^{j-1} x_i\, TNL(X_i))/TNL(X_j) \rfloor$

6.2. Storage of Cell Attribute Values

The values of cell attributes are stored in two cell value array one is cell value array1 C_S and the other is cell value array2 C_μ . The rows of the C_s correspond to the root to leaf path instances of trees in F_r and the column of Cs correspond to the root to leaf path instances of trees in F_c.They are stored in two cell value file, cell value file1 and cell value file2 respectively in row order with one value per logical record. Let < (PR,VR), (PC,VC) > be an ordered set of root to leaf path instances where

- PR=(X_1,X_2,….X_m) is a root to leaf path of TR and θr is omitted from the tree.
- VR= (x_1,x_2,……x_m) is an ordered set of values for the attributes in the path PR, where x_i is an encoded value for Xi, $1 \leq i \leq m$.
- PC= (Y_1,Y_2,…..Y_n) is a root to leaf path of TC and θ_c is omitted from the path.
- VC= (y_1,y_2,….y_n) is an ordered set of values for the attributes in the path PC,where y_i is an encoded value for Y_i,$1 \leq i \leq n$.

The relative number, r1 of the record in the cell value file1 containing the cell value associated with < PR,VR ,(PC,VC) > is obtained as

$$r1=RLN\,((X_1,x_1)\,,\ldots\ldots\,(X_m,x_m))\;NL(\theta_{c1})+CLN((Y_1,y_1)\,,\ldots..\,(Y_m,y_m))+1$$

The corresponding record number r2 in cell value file2 characterizing the fuzziness of record in cell value file1 is given by $r2 = r1$.

Given $r1$ the relative number of a record in the cell value file1, the ordered set of root to leaf path instances < PR,VR ,(PC,VC)> corresponding to $r1$ can be obtained as

(1) First, we obtain the row number for (PR, VR) and then column number for (PC,VC)
NR= [($r1$−1)$NL(\theta c1)$]
NC = r1 – 1 – NR (NL ($\theta c1$))
(2) Then, we obtain the root to leaf path instances corresponding to NR and NC as discussed in section 6.1.

7. ALGORITHM

In this section, we develop the algorithm for creating fuzzy statistical table. The following variables and data types have been used in explaining the algorithmic details.

UD Universe of Discourse, two dimensional string array for domain of attributes of row header and column header of Fuzzy Statistical Table
I Two dimensional string array for Instances of attributes of row header and column header of Fuzzy Statistical Table
MM Two dimensional integer array(column dimension 4) for storing the variables in the mathematical model of Instances of I
TR Binary tree created by making the root nodes in Fr as immediate descendent of a dummy attribute rowroot
TC Binary tree created by making the root nodes in Fc as immediate descendent of dummy attributes columnroot
PR Array to store the nodes of tree TR in row header of Fuzzy Statistical Table
PC Array to store the nodes of tree TC in column header of Fuzzy Statistical Table
VR ,VR1,VR2,V integer array to store the encoded value of Instances of nodes of tree,TR in row header of Fuzzy Statistical Table
VC ,VC1,VC2,V2 integer array to store the encoded value of Instances of nodes of tree TC in column header of Fuzzy Statistical Table
Cs Two dimensional array to store fuzzy statistics of Fuzzy Statistical Table
$C\mu$ Two dimensional array to store membership degrees of fuzzy statistics of Fuzzy Statistical Table
x,y Integer arrays
RLN ,RLN1,RLN2 Row leaf number
CLN,CLN1,CLN2 Column leaf number
M Boolean variable

Develop the following structures

```
structure  treenode
{    string : Name_of_Attribute
      UD : Universe_of_Discourse
  number : Type /* number is either 1 or 2 for Type-1 and Type-2 resp. */
          I : Instance
     MM : Mathematicalmodel
  treenode *Lchild
  treenode *Rchild
}
structure Fuzzytable
{    string : Name_of_table
        list : Cs
        list : Cμ
   treenode * TR
   treenode * TC
}
structure node
{   number : info
    node * next
}
```

Algorithm for Creating Fuzzy Statistical Table

Create_Fuzzy_Statistical_Table( )

```
1. integer n
2. Fuzzytable table
3. initialize table ← null
4. print " Do you want to create a fuzzy statistical table "
5. print " enter 1 for yes and 0 for no "
6. read n
7. if (n=1) then
8.       Input_Fuzzy_Statistical_Table(&table)
9. else
```

10. print " Good Bye"
11. return

Input_Fuzzy_Statistical_Table(table)

1. integer b,n,VR[ ]
2. treenode *PR[ ]
3. print "Enter name of Fuzzy Statistical Table"
4. read table.Name_of_table
5. print " enter b=1 for creating row tree and b=2 for creating column tree "
6. read b /* variable to store code for the construction of tree TR or TC */
7. If (b=1) then
8. table.TR.Name_of_Attribute⟵"Rowroot" /* initialization of root attribute of row header */
9. table.TR.Universe_of_Discourse⟵null
10. table.TR.Type⟵1
11. table.TR.Instance⟵null
12. table.TR.Mathematicalmodel⟵null
13. print " Enter number of children of dummy root attribute in row header "
14. read n /* n is an integer which can take value either 1 or 2 */
15. print " Enter data for row tree "
16. table.TR.Lchild⟵ Create_tree( )
17. If (n=2) then
18. table.TR.Rchild⟵ Create_tree( )
19. If (b=2) then
20. table.TC.Name_of_Attribute⟵"Columnroot" /* initialization of root attribute of Column header */
21. table.TC.Universe_of_Discourse⟵null
22. table.TC.Type⟵1
23. table.TC.Instance⟵null
24. table. TC.Mathematicalmodel⟵null
25. print " Enter number of children of dummy root attribute in column header"
26. read n /* n is an integer which can take value either 1 or 2 */
27. table.TC.Lchild⟵ Create_tree( )
28. If (n=2) then
29. table.TC.Rchild⟵ Create_tree( )
30. initialize PR←null
31. initialize VR← -1
32. If (table.TR.Name_of_Attribute="rowroot") then
33. PR[0]⟵table.TR
34. else
35. search table.TR for table.TR.Name_of_Attribute="rowroot"
36. PR[0]⟵table.TR
37. initialize table. Cs←table.$C\mu$← null
38. CalculatePRVR(&table,1,PR[0].Lchild,&PR,&VR)
39. return

Create_tree( )

1. treenode *q
2. integer n,len,choice
3. print "Input data for the attributes of Fuzzy Statistical Table "
4. print "Enter the name of attribute--->"
5. read q.Name_of_Attribute
6. print "Enter the Universe of discourse as two dimensional string array--->"
7. read q.Universe_of_Discourse

```
print “Enter the type of fuzzy data--->”
read q.Type
print “ Enter the Instances as two dimensional string array--->”
read q.Instance
len←length(q.Instance)
if (q.type=1) then
     print “membership degrees are to be entered by user if yes enter 1 else 0”
     read choice
     if (choice=1) then
         for i ← 0 to (len-1)
                 print “Enter membership degree of q.Instance[i]--->”
                 read q.Mathematicalmodel[i]
      else
            print “Enter data for the mathematical model of q.Instance[i]--->”
             print “1− 1 sign x−breakvalue /denominator −1"
             print “If compliment of some mathematical function enter 1 else 0--->”
             read q.Mathematicalmodel[1][0]
             print “Enter sign used in mathematical model--->”
             read q.Mathematicalmodel[1][1]
             print “ Enter the break value used in mathematical model--->”
             read q.Mathematicalmodel[1][2]
             print “ Enter the denominator value used in mathematical model --->“
             read q.Mathematicalmodel[1][3]
if (q.type=2) then
   for i←0 to (len-1)
        if (check_linguistic_term(q.Instance[i])=true) then
             print “Enter data for the mathematical model of q.Instance[i]--->”
             print “1− 1 sign x−breakvalue /denominator −1"
             print “If compliment of some mathematical model enter 1 else 0--->”
             read q.Mathematicalmodel[i][0]
             print “Enter sign used in mathematical model--->”
             read q.Mathematicalmodel[i][1]
             print “ Enter the break value used in mathematical model--->”
             read q.Mathematicalmodel[i][2]
             print “ Enter the denominator value used in mathematical model --->“
             read q.Mathematicalmodel[i][3]
       else
            print “Enter membership degree of q.Instance[i]--->”
            read q.Mathematicalmodel[i]
print “Enter number of children”
read n
If (n=0) then
     q.Lchild⟵null
     q.Rchild⟵null
     return q
If (n=1) then
    q.Lchild ⟵ Create_tree( )
    q.Rchild⟵null
    return q
If (n=2) then
   q.Lchild⟵ Create_tree( )
```

```
    q.Rchild⟵ Create_tree( )
return q
```

CalculatePRVR (table,index1,q,PR,VR) /* index1 is variable to indexed through PR and VR */

```
treenode *P[ ]
integer j,m,VC[ ],RLN,index2
If (q≠null) then
    PR[index1]⟵q
    for i ⟵ 0 to [length(q.Instance)-1]
       VR[index1]⟵i
        print " Enter data for",q.Name_of_Attribute,"Instance encoded value",i
        If (q.Lchild=null) and (q.Rchild=null) then
          RLN⟵0
          for j ⟵1 to index1 /*calculates row leaf number*/
              m ← NL PR j−1 ,PR j +VR j ∗TNL PR j
              RLN⟵RLN + m
          initialize PC⟵ null
          initialize VC⟵ -1
          If (table.TC.Name_of_Attribute="columnroot") then
              PC[0]⟵table.TC
          else
             search tree table.TC for table.TC.Name_of_Attribute="columnroot"
             PC[0]⟵table.TC
       index2⟵1 /* variable to indexed through PC and VC */
       CalculatePCVC(&table,index1,&PR,&VR ,index2,PC[0].Lchild,&PC,&VC,RLN)
  index1⟵index1+1
  CalculatePRVR(&table,index1,q.Lchild,&PR,&VR)
  CalculatePRVR(&table,index1,q.Rchild,&PR,&VR)
   index1⟵index1-1
VR[index1]⟵ -1
return
```

CalculatePCVC(table,index1,PR,VR,index2,r,PC,VC,RLN)

```
integer i,CLN,j,m
If (r≠null) then
       PC[index2]⟵r
        for i ⟵ 0 to [length(r.Instance)-1]
           VC[index2]⟵i
           print " Enter data for",r.Name_of_Attribute,"Instance encoded value",i
           If (r.Lchild=null) and (r.Rchild=null) then
             CLN⟵0
             for j ⟵1 to index2 /*calculates column leaf number */
                  m ← NL PC j−1 ,PC j +VC j ∗TNL PC j
                  CLN⟵CLN + m
                  print "Enter statistical data--->"
                  read table.Cs[RLN,CLN]
                  print "Enter corresponding membership degree--->"
                  read table.Cμ[RLN,CLN]
           index2⟵index2+1
             CalculatePCVC(&table,index1,&PR,&VR, index2,r.Lchild,&PC,&VC,RLN)
             CalculatePCVC(&table,index1,&PR,&VR, index2,r.Rchild,&PC,&VC,RLN)
             index2⟵index2-1
      VR[index2]⟵ -1
Return
```

8. CONCLUSION

Although there has been a flurry of research activity in database during recent years, it does not cover the issue of storage of linguistic terms or ambiguities in statistical data values. In this paper, we have defined the concept of fuzzy count and addressed the issue of representing fuzziness in statistical environment giving rise to fuzzy statistical database. It alleviates the user from tedious task of computing fuzzy statistics again and again from fuzzy relation. Fuzzy statistical database would be quite useful when used as a decision in areas such as management, decision making and census data evaluation. A logical view for representation of fuzzy statistical database, fuzzy statistical table is discussed. Depending upon the type of category attributes in fuzzy statistical table type-1 and type-2 fuzzy statistical tables are explored. The physical organization technique of fuzzy statistical table is discussed and algorithm for its implementation is also developed. For retrieving the information about a particular tuple, we are developing a query system where DDL and DML are taken care of so that our desired information can be displayed.

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-2 fuzzy relation Employee in a sample of a population

Name	State	Sex	Experience	Salary	Income tax	μ
Sunil	Delhi	M	Little	Low	Low	0.7
Pranav	Delhi	M	Little	Low	Low	0.5
Vimal	Delhi	M	Little	Low	Low	0.5
Samar	Delhi	M	Moderate	40,000-60,000	5000	0.33
Raj	Delhi	M	20	High	High	0.77
Sunil	Delhi	M	Little	Low	Low	0.7
Renu	Delhi	F	Moderate	40,000-60,000	15	0.14
Wasim	Delhi	M	20	High	7000	1
Hiran	Bombay	M	Little	Low	Low	0.08
Dev	Bombay	M	Moderate	High	4000	0.25
Kartikey	Bombay	M	Little	Low	2000	0.03
Dolly	Bombay	F	Moderate	40,000-60,000	High	0.5
Dhruv	Bombay	M	Little	Low	Low	0.08
Anoop	Bombay	M	Little	Low	Low	0.04
Deepak	Bombay	M	Moderate	High	6000	0.33

Table 2. 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 3. An instance of type-2 fuzzy statistical table 2001COUNT in a sample of a population

2001COUNT						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 4. An instance of fuzzy primitive table FS1 in a sample of a population

FS1						Incometax				μ			
						3000	High	Low	4000-7000				
			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 5 An instance of fuzzy primitive table FS2 in a sample of a population

FS2						Incometax				μ			
						3000	High	Low	4000-7000				
			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 with incomplete cross product, FSINC

FSINC	Incometax				μ
	3000	High	Low	4000-7000	

				Exp	10	78	70	60	10	1	0.59	0.63	1
			M		15-20	23	80	50	90	1	0.77	0.8	1
				Sal	30,000	56	43	45	35	1	0.3	0.6	1
	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
			F		15-20	65	35	56	57	1	0.37	0.8	1
				Sal	30,000	85	34	73	54	1	0.3	0.7	1
State					40,000-60,000	45	34	64	57	1	0.26	0.23	1
				Exp	Little	59	86	63	87	0.01	0.08	0.04	0.03
			M		Mod	95	65	56	88	0.25	0.33	0.17	0.2
				Sal	30,000	53	56	36	56	1	0.4	0.41	1
	Bombay	Sex			40,000-60,000	57	44	67	34	1	0.37	0.41	1
				Exp	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
					40,000-60,000	56	67	86	64	1	0.3	0.41	1

Table 7. Encoded value of attributes

Attribute	Instances	Encoded Value
State	Delhi	0
	Bombay	1
Sex	M	0
	F	1
Exp	10	0
	15-20	1
	Little	2
	Moderate	3
Sal	30,000	0
	Low	1
	High	2
	40,000-60,000	3
Incometax	3000	0
	Low	1
	High	2
	4000-7000	3

REFERENCES

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

[2] Zadeh, L.A. (1978),Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1:3-28.

[3] Zadeh, L.A. (1979), A theory of approximate reasoning. In Machine Intelligence 9, J. Hayes, et al., Eds. Ellis Horwood Ltd., Sussex, UK,149-194.

[4] Zadeh, L.A.(1981), PRUF-a meaning representation language for natural language.In Fuzzy Reasoning and itsApplications, E. H. Mamdani and B. R. Gaines, Eds. Academic Press, New York,l-66.

[5] Codd, E.F. (1970), A relational model of data for large shared data banks, ACM, vol.13,no.6.

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

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

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

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

[10] 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.

[11] Paul Chan 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.

[12] 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.

[13] Ozsoyoglu G., Ozsoyoglu Z.M.(1983),An extension of relational algebra for summary tables ,In Proc. Second LBL Workshop on Statistical Database Management.

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

[15] 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.

[16] Ozsoyoglu G. , Ozsoyoglu Z.M.(1982), STBE-A database query language for manipulating summary data,Dep. Comput. Eng. and Sci.,Case Reserve Univ. Cleveland,OH,Tech. Rep. CES-82-8;also In Proc. IEEE COMPDEC Conf.,1984.

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

[18] Ozsoyoglu G. , Ozsoyoglu Z.M.(1985), Statistical database query languages, IEEE Trans.Software Eng.,vol. SE 11,no.10.

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

[20] Raju K.V.S.V.N., Majumdar A.K.(1987), "The study of joins in fuzzy relational databases, Fuzzy sets and systems", 21, 19-34.

[21] Raju K.V.S.V.N., Majumdar A.K.(1988), "Fuzzy functional dependencies and lossless join decomposition of fuzzy relational databases", ACM Transactions Database systems, 13(2),129-166.

[22] Michalewicz Z.(1991),Statistical and Scientific Databases,Z.Michalewicz (Ed.), Ellis Horwood, New York.

[23] 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.

[24] Malvestuto F.M.(1993),A universal scheme approach to statistical databases containing homogeneous summary tables, ACM Transactions on Data systems,vol.18, no.4,pp.678-708.

[25] Paul Kersten R.(1995),The Fuzzy median and the Fuzzy Mad ,In Proceedings of ISUMA- NAFIPS‟95.

[26] Petry F.E.(1996),Fuzzy Databases Principles and Application, Kluwer Academic Publishers, USA.

[27] Klir G.J. ,Bo Yuan(1997),Fuzzy sets and fuzzy logic ,PHI,New York.

[28] Shoshani A.(1997), OLAP and statistical databases:similarities and differences,In Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems,pp. 185 – 196.

[29] Abdullah M.L.,Wong Abdullah, Abu Osman(2004), Fuzzy Sets in the social Sciences: An Overview of Related Researches, Jurnal Teknologi,41(E)Dis.43-54.

[30] James J. Buckley(2004),Fuzzy Statistics, Springer-Verlag Berlin Heidelberg.

[31] Wu Berlin ,Hung T. Nguyen(2006), Fundamentals of Statistics with Fuzzy Data, Springer-Verlag New York.

[32] Casillas J., Sanchez L.(2006),Knowledge extraction from fuzzy data for estimating consumer behavior models IEEE International Conference on Fuzzy System,16-21.

[33] Galindo J., Piattini M., Urrutia A. (2006),Fuzzy databases modeling,design,implementation,Idea Group Publishing. [36] Antova, L., Jansen, T., Koch, C., Olteanu, D. (2008)," Fast and simple processing of uncertain data", In: Proc. of ICDE 2008, pp. 983–992.

[34] Gottwald S.(1980). "A note on fuzzy cardinals." Kybernetika, vol. 16. pp. 156-158.

[35] Benjelloun, O., Das Sarma, A., Halevy, A., Widom, J.(2006) ULDBs," Databases with uncertainty and lineage.", In: Proc. VLDB 2006, pp. 953–964.

[36] Dalvi, N., Suciu, D.: Management of probabilistic data: Foundations and challenges. In: Proc. of PODS 2007, pp. 1–12 (2007).

[37] 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.

[38] 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. [41] Green T.J., Tannen V. (2006), "Models for incomplete and probabilistic information",IEEE Data Eng. Bull. 29, 17–24.

[39] R´e C., Dalvi N., Suciu D.(2006), Query evaluation on probabilistic databases. IEEE Data Eng. Bull. 29, 25–31.

[40] Blanchard N., Cardinal and ordinal theories about fuzzy sets, in: M.M. Gupta and E. Sanchez, Eds. Fuzzy Information and Decision Processes (North-Holland, Amsterdam, 1982) 149 157.

[41] DeLuca A. and Termini S.(1972), A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory, Inform. and Control 20, 301 312.

[42] Dubois D. and Prade H.(1985), Fuzzy cardinality and the modeling of imprecise quantification, Fuzzy Sets and Systems 16,199-230.

[43] Ralescu D. (1995), Testing of fuzzy hypotheses, submitted.

[44] Zadeh L.A.(1983), A computational approach to fuzzy quantifiers in natural languages, Comput. Math. 9,149-184.

[45] Casasnovas J. and Torrens J(2003), Scalar cardinalities of finite fuzzy sets for t-norms and t- conorms, Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems. vol. 11. pp. 599-614.

[46] Kaufmann A.(1977). Introduction a la Theorie des Sous-Ensembles Flous. Masson. Paris..

[47] Wygralak M.(1986), "Fuzzy cardinals based on the generalized equality of fuzzy subsets." Fuzzy Sets and Systems. vol. 18. pp. 338-353.

[48] Delgado M.,Sanchez. D.,Martin-Bautista. M. and Vila M.(2002). "A probabilistic definition of a nonconvex fuzzy cardinality." Fuzzy Sets and Systems. vol. 126. pp. 177-190.

[49] Dubois D. and Prade H.(1985), "Fuzzy cardinality and the modelling of imprecise quantification." Fuzzy Sets and Systems. vol. 16. pp. 199-230