A Hybrid Relational Database Model for Uncertain and Imprecise Information

Hoa Nguyen[1, 2, *]

[1] Information Technology Faculty
Saigon University, Ho Chi Minh City, Vietnam

[2] Information Technology Faculty, Industrial University of Ho Chi Minh City
Ho Chi Minh City, Vietnam

[*]*Corresponding author's email: nguyenhoa [AT] sgu.edu.vn*

ABSTRACT— ***Recent years, many fuzzy or probabilistic database models have been built for representing and handling imprecise or uncertain information of objects in real-world applications. However, relational database models combining the relevance and strength of both fuzzy set and probability theories have rarely been proposed. This paper introduces a new relational database model, as a hybrid one combining consistently fuzzy set theory and probability theory for modeling and manipulating uncertain and imprecise information, where the uncertainty and imprecision of a relational attribute value are represented by a fuzzy probabilistic triple, the computation and combination of relational attribute values are implemented by using the probabilistic interpretation of binary relations on fuzzy sets, and the elimination of redundant data is dealt with by coalescing ε-equivalent tuples. The basic concepts of the classical relational database model are extended in this new model. Then the relational algebraic operations are formally defined accordingly. A set of the properties of the relational algebraic operations is also formulated and proven.***

Keywords— Fuzzy probabilistic triple, ε- equivalence, uncertain and imprecise information, relational algebraic operation

---

1. INTRODUCTION

As we know, the information that we have about the real world may be uncertain and imprecise. Although the classical database models, including the relational database model and object-oriented database model, are useful for modeling, designing, and implementing large-scale systems, they are restricted in representation and handling of uncertain and imprecise information ([3], [5]). For example, the classical relational database model cannot deal with queries such as "find all patients whose daily treatment costs are *high*"; nor "find all players who are 90-95% likely to be the top scorers of the English Premier League in the year 2020", etc., where "*high*" is an imprecision notion ([11], [26]) and "90-95% likely" expresses an uncertainty degree ([9]).

Up to now, there have been many non-classical relational database models researched and developed to overcome the limitations of the classical database models. Some models (e.g. [7], [8], [10], [12], [16], [19], [20], [22], [27], [28]) using only the probability theory could represent and handle uncertain information but not imprecise information of objects. Some other models (e.g. [11], [13], [14], [15], [21], [23]) using only the fuzzy set theory could express and manipulate imprecise information but not uncertain information of objects.

In reality, information may contain both uncertainty and imprecision. For example, the query "find all patients who are *old* and at least 80% likely catch a lung cancer or tuberculosis" contains both imprecise and uncertain information. In such a case, the above-cited models cannot be applied. However, relational database models combining both the fuzzy set theory and probability theory for modeling objects involving both uncertain and imprecise information are rare.

In [24], the authors proposed a fuzzy probabilistic relational database model to represent and deal with uncertain and imprecise information of objects in real-world applications. In this model, each attribute of a tuple in a relation was assigned to a precise value with a probability inferred from the possibility distribution of probability values associated with the tuple. In other words, each tuple in a relation was associated with a fuzzy number as a possibility distribution on the interval [0, 1] representing the aggregated probability for the single value that each attribute in the tuple of the relation can take. Also, in this model, the authors defined the notion of the equivalence of tuples to combine them in relational algebraic operations as the projection, intersection, union and difference. However, in the real world, there are situations in which we do not know exactly the value of each attribute, although we know that the attribute may take one of the values, which can be vague, in a certain set.

In [16], the authors introduced a probabilistic relational database model named PRDB. It was able to represent situations in which we do not know exactly the value of each attribute, but we know the probability interval for it taking one of the values in a candidate set. It means that the model could overcome the shortcoming of the model in [24]. However, the PRDB model could not express and deal with vague information. In [25], the authors extended the model in [24] by allowing relational attributes could take set or tuple type values. However, like the model in [16], this model could not represent and deal with vague information. In [6], the authors proposed a model, where the relational attributes could take a fuzzy set value and each tuple had a probability interval to belong to a relation. The notion of the equivalence of tuples were also defined in this model. However, in [6], only the selection operation was defined while all other algebraic operations are missing.

In [17], the authors extended the PRDB model in [16] with fuzzy set values, resulting in a fuzzy probabilistic relational database model, called FPRDB, that could represent and handle both uncertain and vague information. Nevertheless, in [17], only the selection operation was built while all other algebraic operations are missing.

The model in [18] was an extension of the model in [17] with a complete set of basic fuzzy probabilistic relational algebraic operations, however, in [18], the keys of relational schemas were defined to be certain and precise values. These led to the inconsistency in definitions of some relational algebraic operations such as the operations of the projection, intersection, union and difference of relations and caused information loss for objects in a database. In this paper, using the concept of fuzzy probabilistic triples introduced in [4] and the probabilistic interpretation of binary relations on fuzzy sets presented in [17], we propose an uncertain and fuzzy relational database model, denoted UFRDB, as a hybrid one combining probability theory and fuzzy set theory, where the notion of the uncertainty and imprecision of attribute values, and of the equivalence of relational tuples are defined as the basis to build UFRDB algebraic operations and overcome the shortcoming of the model in [18] for representing and manipulating both imprecise and uncertain information in practice.

The mathematical basis for UFRDB is summarized in Section 2. The data model of UFRDB are presented in Section 3. Sections 4 and 5 respectively present the UFRDB algebraic operations and their properties. Finally, Section 6 concludes the paper and outlines further research directions.

2. BASIC PROBABILITY AND FUZZY SET DEFINITIONS

This section presents some notions of probability and fuzzy sets as a mathematical foundation for extending the classical relational database model (CRDB) to the UFRDB model.

2.1 Mass Assignment and Probability of Fuzzy Binary Relations

For a probabilistic interpretation of binary relations on fuzzy sets, the mass assignment based on the voting model of fuzzy sets in [1] and [2] is defined as follows.

Definition 1. Let $A = \sum_{i=1,\ n} \sum_{j=1,\ m_i} x_{i,j} : y_i$ be a normal fuzzy set on a domain U, where n, $m_i \in N$, and $y_i > y_j$ if $i < j$, $\forall i =1,\dots, n$ and $\forall j =1,\dots, m_i$. The *mass assignment* corresponding to A is a mapping m_A: $2^U \rightarrow [0, 1]$ that is defined by $m_A(z_1) = y_1 – y_2, \dots, m_A(\cup_{i=1,\ j} z_i) = y_j – y_{j+1}, \dots, m_A(\cup_{i=1,\ n} z_i) = y_n$, where $z_i = \cup_{j=1,\ m_i}\{x_{i,j}\}$.

We note that, the mass assignment $m_A(z_1) = y_1 – y_2, \dots, m_A(\cup_{i=1,\ j} z_i) = y_j – y_{j+1}, \dots, m_A(\cup_{i=1,\ n} z_i) = y_n$ can be denoted by $m_A = z_1: y_1 – y_2, \dots, \cup_{i=1,\ j} z_i: y_j – y_{j+1}, \dots, \cup_{i=1,\ n} z_i: y_n$.

As in ([4], [17]), the probabilistic interpretation of a binary relation on fuzzy sets is then defined as the probability for the relation being true as below.

Definition 2. Let A be a fuzzy set on a domain U, B be a fuzzy set on a domain V, and θ be a binary relation from $\{=, \neq, \leq, \geq, <, >\}$ assumed to be valid on $(U \times V)$. The *probabilistic interpretation* of the relation $A\ \theta\ B$, denoted by $prob(A\ \theta\ B)$, is a value in [0, 1] that is defined by $\sum_{S \subseteq U,\ T \subseteq V} p(u\ \theta\ v|\ u \in S,\ v \in T).m_A(S).m_B(T)$, where m_A, m_B are the mass assignments corresponding to A and B, respectively, and $p(u\ \theta\ v|\ u \in S,\ v \in T)$ is the conditional probability of $u\ \theta\ v$ given $u \in S$ and $v \in T$.

Intuitively, given fuzzy propositions "*x* is *A*" and "*y* is *B*", $prob(A\ \theta\ B)$ is the probability for $x\ \theta\ y$ being true.

Definition 3. Let A and B be two fuzzy sets on a domain U. The *probabilistic interpretation* of the relation $A \Rightarrow B$, denoted by $prob(A \Rightarrow B)$, is a value in [0, 1] that is defined by $\sum_{S,T \subseteq U} p(u \in T|\ u \in S).m_A(S).m_B(T)$, where m_A, m_B are the mass assignments corresponding to A and B, respectively, and $p(u \in T|\ u \in S)$ is the conditional probability for $u \in T$ given $u \in S$.

The intuitive meaning of $prob(A \Rightarrow B)$ is that, given a fuzzy proposition "$x \in A$", $prob(A \Rightarrow B)$ is the probability for $x \in B$ being true. In other words, it is the fuzzy conditional probability of $x \in B$ given $x \in A$.

Example 1. Let *about_6* = {5: 0.9, 6: 1.0, 7: 0.9} and *about_7* = {6: 0.9, 7: 1.0, 8: 0.9} be the fuzzy sets on the real number set R, then the mass assignments corresponding to *about_6* and *about_7* are m_{about_6} ={6}:0.1, {5, 6, 7}:0.9 and m_{about_7} = {7}:0.1, {6, 7, 8}:0.9. The probabilistic interpretation of *about_7* ⇒ *about_6* is computed as follows:

$$prob(about\_7 \Rightarrow about\_6) = p(u\in\{6\} \mid u\in\{7\}).m_{about\_7}(\{7\}).m_{about\_6}(\{6\}) + p(u\in\{6\} \mid u \in\{6,7,8\}).m_{about\_7}(\{6,7,8\}).m_{about\_6}(\{6\}) + p(u\in\{5,6,7\} \mid u \in\{7\}).m_{about\_7}(\{7\}).m_{about\_6}(\{5,6,7\}) + p(u\in\{5,6,7\} \mid u\in\{6,7,8\}).m_{about\_7}(\{6,7,8\}).m_{about\_6}(\{5,6,7\})$$
$$= 0\times 0.1\times0.1 + 1/3\times0.9\times0.1 + 1\times0.1\times0.9 + 2/3\times0.9\times0.9 = 0.66.$$

2.2 Combination Strategies of Probability Intervals

Let two events e_1 and e_2 have probabilities in the intervals $[L_1, U_1]$ and $[L_2, U_2]$, respectively, then the probability intervals of the conjunction event $e_1 \wedge e_2$, disjunction event $e_1 \vee e_2$, and difference event $e_1 \wedge \neg e_2$ can be computed by alternative strategies. In this work, we employ the conjunction, disjunction, and difference strategies given in [9] as presented in Table 1, where ⊗, ⊕, and ⊖ denote the conjunction, disjunction, and difference operators, respectively.

Table 1: Definitions of probabilistic combination strategies

Strategy	Operators
Ignorance	$([L_1, U_1]\otimes_{ig}[L_2, U_2]) = [max(0, L_1 + L_2 - 1), min(U_1, U_2)]$ $([L_1, U_1]\oplus_{ig}[L_2, U_2]) = [max(L_1, L_2), min(1, U_1 + U_2)]$ $([L_1, U_1]\ominus_{ig}[L_2, U_2]) = [max(0, L_1 - U_2), min(U_1, 1 - L_2)]$
Independence	$([L_1, U_1]\otimes_{in}[L_2, U_2]) = [L_1 . L_2, U_1 . U_2]$ $([L_1, U_1]\oplus_{in}[L_2, U_2]) = [L_1+L_2 - (L_1.L_2), U_1+U_2 - (U_1.U_2)]$ $([L_1, U_1]\ominus_{in}[L_2, U_2]) = [L_1 . (1 - U_2), U_1 . (1 - L_2)]$
Positive correlation	$([L_1, U_1]\otimes_{pc}[L_2, U_2]) = [min(L_1, L_2), min(U_1, U_2)]$ $([L_1, U_1]\oplus_{pc}[L_2, U_2]) = [max(L_1, L_2), max(U_1, U_2)]$ $([L_1, U_1]\ominus_{pc}[L_2, U_2]) = [max(0, L_1 - U_2), max(0, U_1 - L_2)]$
Mutual exclusion	$([L_1, U_1]\otimes_{me}[L_2, U_2]) = [0, 0]$ $([L_1, U_1]\oplus_{me}[L_2, U_2]) = [min(1, L_1 + L_2), min(1, U_1 + U_2)]$ $([L_1, U_1]\ominus_{me}[L_2, U_2]) = [L_1, min(U_1, 1 - L_2)]$

In the following sections, the notation $[L_1, U_1] \leq [L_2, U_2]$ is used to denote $L_1 \leq L_2$ and $U_1 \leq U_2$ whereas the notation $[L_1, U_1] \subseteq [L_2, U_2]$ is for $L_2 \leq L_1$ and $U_1 \leq U_2$. Also, a single probability value ε can be treated as the probability interval [ε, ε].

2.3 Combination Strategies of Fuzzy Probabilistic Triples

For representing imprecise and uncertain attribute values in UFRDB, we use the notion of fuzzy probabilistic triples in [4] and [17] extended from probabilistic triples in [9] and is defined as below.

Definition 4. Let $\mathcal{X}$ be a non-empty set. A *fuzzy probabilistic triple* on $\mathcal{X}$ is defined to be of the form $\langle V, \alpha, \beta\rangle$, where V is a finite subset of $\mathcal{X}$, and α and β are respectively lower and upper bound probability distributions on V.

In UFRDB, the sets $\mathcal{X}$ and V can consist of fuzzy set values. Informally, a fuzzy probabilistic triple $\langle V, \alpha, \beta\rangle$ assigns to each element $x \in V$ a probability $p(x)$ where $\alpha(x) \leq p(x) \leq \beta(x)$ to represent the uncertainty degree that an object may take the value x in V, which can be an imprecise value.

Example 2. Suppose the daily treatment cost of a patient is estimated within about 6 or 7 (USD) with a probability for each between 0.4 and 0.6. Then this information can be represented by the fuzzy probabilistic triple $\langle V, \alpha, \beta\rangle$ = ⟨{*about_6*, *about_7*}, 0.8*u*, 1.2*u*⟩, where *about_6* and *about_7* are fuzzy sets given as in Example 1, defining the imprecise treatment costs of the patient and u is the uniform distribution over V = {*about_6*, *about_7*}. Here, 0.8u and 1.2u are respectively the probability distributions α and β with $\alpha(x) = 0.8u(x) = 0.8(1/2) = 0.4$ and $\beta(x) = 1.2u(x) = 1.2(1/2) = 0.6$, $\forall x \in V$ = {*about_6*, *about_7*}.

For building UFRDB algebraic operations, we employ combination strategies of fuzzy probabilistic triples in [4]. We note that, here $h(v)$ denotes the height of a fuzzy set v, whereby v is a normal fuzzy set if and only if $h(v) = 1$.

Definition 5. Let $fpt_1 = \langle V_1, \alpha_1, \beta_1\rangle$ and $fpt_2 = \langle V_2, \alpha_2, \beta_2\rangle$ be two fuzzy probabilistic triples, and ⊗ be a probabilistic conjunction strategy. Then the *conjunction* of fpt_1 and fpt_2 under ⊗, denoted by $fpt_1 \otimes fpt_2$, is the fuzzy probabilistic triple $fpt = \langle V, \alpha, \beta\rangle$, such that:

1. $V = \{v = v_1 \cap v_2 \mid v_1\in V_1, v_2\in V_2, h(v) = 1, [\alpha_1(v_1), \beta_1(v_1)] \otimes [\alpha_2(v_2), \beta_2(v_2)] \neq [0,0]\}$, and
2. $[\alpha(v), \beta(v)] = \oplus_{me:\ v_1\in V_1, v_2\in V_2, v = v_1\cap v_2} [\alpha_1(v_1), \beta_1(v_1)] \otimes [\alpha_2(v_2), \beta_2(v_2)]$, for every $v\in V$, where $\oplus_{me}$ is the mutual exclusion probabilistic disjunction strategy.

We note that, unlike the combination strategies of probabilistic triples in [9], where each v_1 and v_2 in V_1 and V_2 respectively is elementary and non-fuzzy, here v_1 and v_2 may be fuzzy sets, since there can be more than one pair $(v_1, v_2) \in V_1 \times V_2$ such that $v = v_1 \cap v_2$. So the probability intervals for those pairs must be combined using the mutual exclusion probabilistic disjunction strategy $\oplus_{me}$ in the above computation of $[\alpha(v), \beta(v)]$.

Example 3. Let $fpt_1 = \langle\{about\_40, about\_50\}, u, u\rangle$ and $fpt_2 = \langle\{about\_50, about\_60\}, 0.8u, 1.2u\rangle$ be fuzzy probabilistic triples, where *about_40*, *about_50* and *about_60* are fuzzy sets with $h(about\_40 \cap about\_60) < 1$, then $fpt_1 \otimes_{in} fpt_2$ with the independence probabilistic conjunction strategy is the fuzzy probabilistic triple $fpt = \langle\{about\_50\}, 0.2u, 0.3u\rangle$.

Next, the disjunction and difference of fuzzy probabilistic triples in turn are defined as below.

Definition 6. Let $fpt_1 = \langle V_1, \alpha_1, \beta_1\rangle$ and $fpt_2 = \langle V_2, \alpha_2, \beta_2\rangle$ be two fuzzy probabilistic triples, and $\oplus$ be a probabilistic disjunction strategy. Then the *disjunction* of fpt_1 and fpt_2 under $\oplus$, denoted by $fpt_1 \oplus fpt_2$, is the fuzzy probabilistic triple $fpt = \langle V, \alpha, \beta\rangle$, such that:

1. $V = P \cup Q \cup R$, where $P = \{v_1 \in V_1 \mid \neg\exists v_2 \in V_2: h(v_1 \cap v_2) = 1\}$, $Q = \{v_2 \in V_2 \mid \neg\exists v_1 \in V_1: h(v_1 \cap v_2) = 1\}$, and $R = \{v_1 \cap v_2 \mid v_1 \in V_1, v_2 \in V_2, h(v_1 \cap v_2) = 1\}$, and
2. $[\alpha(v), \beta(v)] = \begin{cases} [\alpha_1(v), \beta_1(v)], \forall v \in P \\ [\alpha_2(v), \beta_2(v)], \forall v \in Q \\ \oplus_{me: v_1 \in V_1, v_2 \in V_2, v = v_1 \cap v_2}[\alpha_1(v_1), \beta_1(v_1)] \oplus [\alpha_2(v_2), \beta_2(v_2)], \forall v \in R. \end{cases}$

Definition 7. Let $fpt_1 = \langle V_1, \alpha_1, \beta_1\rangle$ and $fpt_2 = \langle V_2, \alpha_2, \beta_2\rangle$ be two fuzzy probabilistic triples, and $\ominus$ be a probabilistic difference strategy. Then the *difference* of fpt_1 and fpt_2 under $\ominus$, denoted by $fpt_1 \ominus fpt_2$, is the fuzzy probabilistic triple $fpt = \langle V, \alpha, \beta\rangle$, such that:

1. $V = P \cup Q$, where $P = \{v_1 \in V_1 \mid \neg\exists v_2 \in V_2: h(v_1 \cap v_2) = 1\}$, $Q = \{v = v_1 \cap v_2 \mid v_1 \in V_1, v_2 \in V_2, h(v_1 \cap v_2) = 1$ and $[\alpha_1(v_1), \beta_1(v_1)] \ominus [\alpha_2(v_2), \beta_2(v_2)] \neq [0, 0]\}$, and
2. $[\alpha(v), \beta(v)] = \begin{cases} [\alpha_1(v), \beta_1(v)], \forall v \in P \\ \oplus_{me: v_1 \in V_1, v_2 \in V_2, v = v_1 \cap v_2}[\alpha_1(v_1), \beta_1(v_1)] \ominus [\alpha_2(v_2), \beta_2(v_2)], \forall v \in Q. \end{cases}$

3. PROPOSED UFRDB MODEL

As for CRDB, the schema and relation are the fundamental concepts in the UFRDB model.

3.1 UFRDB Schemas

A UFRDB schema describes a set of relational attributes and their associated sets of fuzzy probabilistic triples representing possible values of objects in UFRDB, as defined below.

Definition 8. A *UFRDB schema* is a pair $R = (\boldsymbol{U}, \wp)$, where
1. $\boldsymbol{U} = \{A_1, A_2, \dots, A_k\}$ is a set of pairwise different attributes.
2. $\wp$ is a function that maps each attribute $A \in \boldsymbol{U}$ to the set of all fuzzy probabilistic triples on the value domain of A.

Note that as in CRDB ([3], [5]), for simplicity, the notation $R(\boldsymbol{U}, \wp)$ and then R can be used to denote $R = (\boldsymbol{U}, \wp)$. In addition, the value domain of each attribute A is denoted by $dom(A)$.

3.2 UFRDB Relations

A UFRDB relation is an instance of a UFRDB schema in which each attribute may take imprecise and uncertain values represented by a fuzzy probabilistic triple as in the following definition.

Definition 9. Let $\boldsymbol{U} = \{A_1, A_2, \dots, A_k\}$ be a set of k pairwise different attributes. A *UFRDB relation r* over the schema $R(\boldsymbol{U}, \wp)$ is a finite set $\{t \mid t = (\langle V_1, \alpha_1, \beta_1\rangle, \langle V_2, \alpha_2, \beta_2\rangle, \dots, \langle V_k, \alpha_k, \beta_k\rangle)\}$, in which each element t is a list of k fuzzy probabilistic triples such that $\langle V_i, \alpha_i, \beta_i\rangle$ belongs to the set $\wp(A_i)$ and $V_i \neq \emptyset$, for every $i = 1, 2, \dots, k$.

Each element t in the relation r over $R(\boldsymbol{U}, \wp)$ is called a *tuple* on $\boldsymbol{U}$. For each tuple t, the fuzzy probabilistic triple $\langle V_i, \alpha_i, \beta_i\rangle$ represents the imprecise and uncertain value of the attribute A_i of the tuple t. We write $t.A_i$ to denote $\langle V_i, \alpha_i, \beta_i\rangle$. For each subset of attributes $X \subseteq \{A_1, A_2, \dots, A_k\}$, the notation $t[X]$ is used to denote the rest of t after eliminating the fuzzy probabilistic triples of those attributes that do not belong to X.

As in [3], [6] and [8], our UFRDB adopts the *closed world assumption*. It means that, for each attribute A_i and $v \in dom(A_i) - V_i$, the probability for A_i taking v is 0. In addition, each precise (or crisp) value $v \in V_i$ is considered as a special fuzzy set on $dom(A_i)$ with the membership function $\mu_v(v) = 1$ and $\mu_v(x) = 0\ \forall x \in dom(A_i)$ and $x \neq v$.

Example 4. Assuming a schema **PATIENT**, where $\boldsymbol{U}$ = {P_ID, P_NAME, P_AGE, P_DISEASE, P_COST}, a simple relation PATIENT (over **PATIENT**) about patients at the clinic of a hospital is shown in Table 2. In the relation, the attributes P_ID, P_NAME, P_AGE, P_DISEASE, and P_COST respectively describe the information about the identifier, name, age, disease, and daily treatment cost of each patient. In reality, while being diagnosed, the actual disease of a patient may be still uncertain. Similarly, during the treatment process, the daily treatment cost for a patient can be just an estimation. It is noted that, for each attribute $A \in \boldsymbol{U}$ in the schema **PATIENT**($\boldsymbol{U}$, $\wp$), $\wp(A)$ includes all fuzzy probabilistic triples on the value domain of A (Definition 8). In addition, for simplicity, each fuzzy probabilistic triple $\langle V, u, u\rangle$, with $V = \{v\}$ and u is the uniform distribution over V, will be represented as a single value v. Because if an attribute takes such a fuzzy probabilistic triple, then, actually it only takes a value v with the probability is 1 (Definition 4). In other words, the attribute certainly takes the value v.

Table 2: Relation PATIENT

P_ID	P_NAME	P_AGE	P_DISEASE	P_COST
P215	John	$\langle\{65\}, u, u\rangle$	$\langle\{$*lung cancer, tuberculosis*$\}, 0.8u, 1.2u\rangle$	$\langle\{30, 35\}, 0.7u, 1.3u\rangle$
P226	Paul	$\langle\{$*middle_aged, approx_40*$\}, u, u\rangle$	$\langle\{$*hepatitis, cirrhosis*$\}, 0.9u, 1.3u\rangle$	$\langle\{$*about_6, about_7*$\}, 0.8u, 1.2u\rangle$
P238	Ann	$\langle\{$*old*$\}, u, u\rangle$	$\langle\{$*cholecystitis*$\}, u, u\rangle$	$\langle\{8\}, u, u\rangle$
P382	Selena	$\langle\{$*young*$\}, u, u\rangle$	$\langle\{$*bronchitis*$\}, u, u\rangle$	$\langle\{$*about_7*$\}, u, u\rangle$

In real-world applications, fuzzy set values of attributes of the relation PATIENT, such as *about_6*, *about_7*, *approx_40*, *young*, *middle_aged*, and *old*, should be defined compatibly and consistently with the meaning of the information represented by them. For this simple example of Definition 9, one can define *about_6* = {5: 0.9, 6: 1.0, 7: 0.9} and *about_7* = {6: 0.9, 7: 1.0, 8: 0.9} as fuzzy set values representing the likely imprecise daily treatment costs of the patient Paul who has hepatitis or cirrhosis. Similarly, *approx_40* ={39: 0.9, 40: 1.0, 41: 0.9}, and *middle_aged*, *old*, *young* whose membership functions depicted as below could be used as fuzzy set values representing the imprecise ages of the patients Paul, Ann and Selena, respectively.

$$young(x) = \begin{cases} 1, \ \forall x \in \{0, 1, ..., 20\} \\ (35 - x)/15, \ \forall x \in \{21, 22, ..., 34\}, \end{cases} \quad middle\_aged(x) = \begin{cases} (x - 20)/15, \ \forall x \in \{21, 22, ..., 34\} \\ 1, \ \forall x \in \{35, 36, ..., 45\} \\ (60 - x)/15, \ \forall x \in \{46, 47, ..., 59\}, \end{cases}$$

$$old(x) = \begin{cases} (x - 45)/15, \ \forall x \in \{46, 47, ..., 59\} \\ 1, \ \forall x \in \{60, 61, ..., 90\}. \end{cases}$$

Now, the notion of an uncertain and fuzzy relational database is defined as follows.

Definition 10. *An uncertain and fuzzy relational database* over a set of attributes is a set of UFRDB relations corresponding to the set of their UFRDB schemas.

Note that, if we only care about a unique relation over a schema then we can unify its symbol name with its schema's name.

3.3 UFRDB Equivalent Tuples

As we know, the classical relational database model does not allow redundant tuples in a relation, i.e., those whose respective attribute values are equal. For non-classical relational database models, different tuples in a relation, whose respective attribute values are approximately equal, are considered as redundant tuples and should be handled by eliminating or coalescing. Such redundant tuples, in the models [24] and [27], were called value-equivalent tuples, an extended notion of the notion about the equality of tuples in the classical relational database model. For the model in [24], where relational attributes could take only precise values and the uncertain membership degree of tuples was a possibility distribution of probability values, the authors introduced the notion of value-equivalence. Two tuples were said to be value-equivalent if and only if their respective relational attribute values are equal. Then they should be coalesced into a single tuple with the same relational attribute values and the combined uncertain membership degree as the sum of their ones. Similarly, identical tuples as the result of the projection, union, intersection and difference operations were also coalesced.

For the model in [27], where the relational attribute values were precise and the uncertain membership degree of each tuple was a single probability value, the authors added the notion of ε-equality. Two tuples were said to be ε-equal if and only if they are value-equivalent, as defined in [24], and the absolute difference of their probabilistic attribute values is less than ε.

For our UFRDB model, in order to be coherent with its fuzzy probabilistic framework, where relational attribute values can be proper fuzzy sets associated with probability distributions, we introduce a probability measure for two values of the same attribute in two different tuples being equal and evaluate the likelihood of the value equality of two tuples and propose the notion of ε-equivalence as in the following definitions.

Definition 11. Let $R(\boldsymbol{U}, \wp)$ be a UFRDB schema, t_1 and t_2 be two tuples on $\boldsymbol{U}$, A be an attribute of $\boldsymbol{U}$, and ⊗ be a probabilistic conjunction strategy. The *probability interval* for the values of the attribute A of two tuples t_1 and t_2 being equal under ⊗, denoted by $p(t_1.A =_{\otimes} t_2.A)$, is $[\Sigma_{v\in V}\ \alpha(v).prob(v_1 = v_2), min(1, \Sigma_{v\in V}\ \beta(v).prob(v_1 = v_2))]$, where $t_1.A = \langle V_1, \alpha_1, \beta_1\rangle$, $t_2.A = \langle V_2, \alpha_2, \beta_2\rangle$ and $[\alpha(v), \beta(v)] = [\alpha_1(v_1), \beta_1(v_1)] \otimes [\alpha_2(v_2), \beta_2(v_2)]$, $\forall v = (v_1, v_2) \in V = V_1 \times V_2$.

Definition 12. Let $R(\boldsymbol{U}, \wp)$ be a UFRDB schema, t_1 and t_2 be two tuples on $\boldsymbol{U}$, and $\varepsilon\in[0, 1]$. Then t_1 and t_2 are said to be *ε- equivalent* on $\boldsymbol{U}$ with respect to a probabilistic conjunction strategy ⊗, denoted by $t_1 \approx_{\varepsilon\otimes} t_2$, if and only if $\otimes_{A\in \boldsymbol{U}}\ p(t_1.A =_{\otimes} t_2.A) \geq \varepsilon$.

Intuitively, the concept of the ε-equivalence is to coalesce two UFRDB tuples in a relation under some probabilistic combination strategy if their equality likelihood is greater than or equal to a certain threshold ε, or not to coalesce them otherwise. The number ε is called an *equivalent threshold* of tuples on $\boldsymbol{U}$. It is easy to see that the definition of equal tuples in the classical relational database model is a special case of our definition with ε = 1.

Example 5. Let t_1= (*P302*, Mary, ⟨{21}, *u*, *u*⟩, ⟨{*bronchitis*}, *u*, *u*⟩, ⟨{*about_7*}, *u*, *u*⟩) and t_2= (*P302*, Mary, ⟨{21}, *u*, *u*⟩, ⟨{*bronchitis*}, *u*, *u*⟩, ⟨{*about_6*, *about_7*}, 0.8*u*, 1.2*u*⟩) be two tuples on the set of the attributes $\boldsymbol{U}$ = {P_ID, P_NAME, P_AGE, P_DISEASE, P_COST} of the schema **PATIENT** in Example 4, then $\otimes_{in\ A\in \boldsymbol{U}}\ p(t_1.A =_{\otimes in} t_2.A) = [1, 1]_{\otimes in}[1, 1]_{\otimes in}[1, 1]_{\otimes in}\ [1, 1]_{\otimes in}[0.232, 0.348] = [0.232, 0.348]$ under the independence probabilistic conjunction strategy $\otimes_{in}$, where $p(t_1.\text{P\_COST} =_{\otimes in} t_2.\text{P\_COST}) = [0.232, 0.348]$, $p(t_1.A =_{\otimes in} t_2.A) = [1, 1]$, $\forall A\in\boldsymbol{U}, A \neq$ P_COST (Definition 11). So, t_1 and t_2 are equivalent on $\boldsymbol{U}$ under every equivalent threshold $\varepsilon \in [0, 0.232]$ and the independence probabilistic conjunction strategy $\otimes_{in}$.

In the rest of this paper, we implicitly assume that for each UFRDB relation r over a schema $R(\boldsymbol{U}, \wp)$, there exists a number $\varepsilon \in (0, 1]$ such that there are not any two tuples in r being equivalent under the threshold ε (i.e. ε is an equivalent threshold of tuples on $\boldsymbol{U}$).

4. UFRDB ALGEBRAIC OPERATIONS

As for CRDB ([3], [5]), the basic relational algebraic operations on UFRDB are the selection, projection, Cartesian product, join, intersection, union, and difference. We now extend those operations of CRDB for UFRDB taking into account imprecise and uncertain values of relational attributes.

4.1 Selection

For defining the selection operation, we present the formal syntax and semantics of selection conditions by extending those definitions of CRDB with probability and fuzzy set values. We start with the syntax of selection expressions as in the following definition.

Definition 13. Let R be a UFRDB schema and $\mathcal{X}$ be a set of relational tuple variables. Then *selection expressions* are inductively defined and have one of the following forms:

1. $x.A\ \theta\ c$, where $x \in \mathcal{X}$, A is an attribute in R, θ is a binary relation from {=, ≠, ≤, ≥, <, >, ⇒}, and c is a single value or a fuzzy set.
2. $x.A_1 =_{\otimes} x.A_2$, where $x \in \mathcal{X}$, A_1 and A_2 are two different attributes in R, and ⊗ is a probabilistic conjunction strategy.
3. $E_1 \otimes E_2$, where E_1 and E_2 are selection expressions on the same relational tuple variable, and ⊗ is a probabilistic conjunction strategy.
4. $E_1 \oplus E_2$, where E_1 and E_2 are selection expressions on the same relational tuple variable, and ⊕ is a probabilistic disjunction strategy.

Example 6. Consider the schema **PATIENT** in Example 4, the selection of "all patients who get cirrhosis and pay the daily treatment cost of about 6 USD" can be expressed by the selection expression *x*.P_DISEASE = *cirrhosis* ⊗ *x*.P_COST ⇒ *about_6*.

In UFRDB, each selection condition is a logical combination of selection expressions with probability intervals to be satisfied as in the following definition.

Definition 14. Let R be a UFRDB schema. Then *selection conditions* are inductively defined as follows:

1. If E is a selection expression and $[L, U]$ is a subinterval of [0, 1], then $(E)[L, U]$ is a selection condition.
2. If ϕ and ψ are selection conditions on the same tuple variable, then $\neg\phi$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$ are selection conditions.

Example 7. Given the schema **PATIENT** in Example 4, the selection of "all patients who are old with a probability of at least 0.5 and have lung cancer with a probability of at least 0.9" can be done using the selection condition (x.P_AGE $\Rightarrow$ *old*)[0.5, 1.0] $\wedge$ (x.P_DISEASE = *lung cancer*)[0.9, 1.0].

The probabilistic interpretation (i.e., semantics) of selection expressions is defined by extending those definitions of CRDB with the probabilistic combination strategies and binary relations on fuzzy sets as follows.

Definition 15. Let R be a UFRDB schema, r be a relation over R, x be a tuple variable, and t be a tuple in r. The *probabilistic interpretation* of selection expressions with respect to R, r and t, denoted by $prob_{R,r,t}$, is the partial mapping from the set of all selection expressions to the set of all closed subintervals of [0, 1] that is inductively defined as follows:

1. $prob_{R,r,t}(x.A\ \theta\ c) = [\Sigma_{v\in V}\ \alpha(v).prob(v\ \theta\ c),\ min(1,\ \Sigma_{v\in V}\ \beta(v).prob(v\ \theta\ c))]$, where $t.A = \langle V, \alpha, \beta\rangle$.
2. $prob_{R,r,t}(x.A_1 =_{\otimes} x.A_2) = [\Sigma_{v\in V}\ \alpha(v).prob(v_1 = v_2),\ min(1,\ \Sigma_{v\in V}\ \beta(v).prob(v_1 = v_2))]$, where $t.A_1 = \langle V_1, \alpha_1, \beta_1\rangle$, $t.A_2 = \langle V_2, \alpha_2, \beta_2\rangle$ and $[\alpha(v), \beta(v)] = [\alpha_1(v_1), \beta_1(v_1)] \otimes [\alpha_2(v_2), \beta_2(v_2)]$, $\forall v = (v_1, v_2) \in V = V_1{\times}V_2$.
3. $prob_{R,r,t}(E_1 \otimes E_2) = prob_{R,r,t}(E_1) \otimes prob_{R,r,t}(E_2)$.
4. $prob_{R,r,t}(E_1 \oplus E_2) = prob_{R,r,t}(E_1) \oplus prob_{R,r,t}(E_2)$.

Intuitively, $prob_{R,r,t}(x.A\ \theta\ c)$ is the probability interval for the attribute A of the tuple t having a value v such that $v\ \theta\ c$, while $prob_{R,r,t}(x.A_1 =_{\otimes} x.A_2)$ is the probability interval for the attributes A_1 and A_2 of the tuple t having values v_1 and v_2, respectively, such that $v_1 = v_2$.

Example 8. Let R denote the schema **PATIENT** and r denote the relation PATIENT in Example 4. Consider the second tuple in r, denoted by t_2. By Definition 3, one has $prob(about\_6 \Rightarrow about\_6) = 0.94$ and $prob(about\_7 \Rightarrow about\_6) = 0.66$. Consequently, $prob_{R,r,t_2}(x.\text{P\_COST} \Rightarrow about\_6) = [0.8u(about\_6).prob(about\_6 \Rightarrow about\_6) + 0.8u(about\_7).prob(about\_7 \Rightarrow about\_6), min(1, 1.2u(about\_6).prob(about\_6 \Rightarrow about\_6) + 1.2u(about\_7).prob(about\_7 \Rightarrow about\_6))] = [0.8 \times 0.5 \times 0.94 + 0.8 \times 0.5 \times 0.66, min(1, 1.2 \times 0.5 \times 0.94 + 1.2 \times 0.5 \times 0.66)] = [0.64, 0.96]$.

On the basis of the probabilistic interpretation of selection expressions, the satisfaction of selection conditions in UFRDB is defined as below.

Definition 16. Let R be a UFRDB schema, r be a relation over R, and $t \in r$. The *satisfaction* of selection conditions under $prob_{R,r,t}$ is defined as follows:

1. $prob_{R,r,t} \vDash (E)[L,U]$ if and only if (iff) $prob_{R,r,t}(E) \subseteq [L,U]$.
2. $prob_{R,r,t} \vDash \neg\phi$ iff $prob_{R,r,t} \vDash \phi$ does not hold.
3. $prob_{R,r,t} \vDash \phi \wedge \psi$ iff $prob_{R,r,t} \vDash \phi$ and $prob_{R,r,t} \vDash \psi$.
4. $prob_{R,r,t} \vDash \phi \vee \psi$ iff $prob_{R,r,t} \vDash \phi$ or $prob_{R,r,t} \vDash \psi$.

Note that, in the classical relational database model, the concepts of selection expression and selection condition are identical, where probability intervals [L, U] in selection conditions being always equal to [1.0, 1.0]. This also means that the satisfaction of selection conditions in the classical relational database model is a special case of that in UFRDB.

Now, the selection operation on a relation in UFRDB is defined as follows.

Definition 17. Let R be a UFRDB schema, r be a relation over R, and ϕ be a selection condition over a tuple variable x. The *selection* on r with respect to ϕ, denoted by $\sigma_\phi(r)$, is the relation $r^* = \{t \in r \mid prob_{R,r,t} \vDash \phi\}$ over R, including all satisfying tuples of the selection condition ϕ.

Example 9. Consider the relation PATIENT in Example 4. Then, the query "Find all patients who have cirrhosis with a probability between 0.4 and 0.7 and pay the daily treatment cost of about 6 USD with a probability of at least 0.6" can be done by the selection operation σ_ϕ(PATIENT) with ϕ = (x.P_DISEASE = *cirrhosis*)[0.4, 0.7] $\wedge$ (x.P_COST $\Rightarrow$ *about_6*)[0.6, 1.0].

Only the second tuple (*P226*, Paul, $\langle$\{*middle_aged*, *approx_40*\}, u, $u\rangle$, $\langle$\{*hepatitis*, *cirrhosis*\}, 0.9u, 1.3$u\rangle$, $\langle$\{*about_6*, *about_7*\}, 0.8u, 1.2$u\rangle$) in Example 4 satisfies ϕ, because $prob_{R,r,t_2}(x.\text{P\_DISEASE} = cirrhosis) = [0.45, 0.65] \subseteq [0.4, 0.7]$ and by Example 8 $prob_{R,r,t_2}(x.\text{P\_COST} \Rightarrow about\_6) = [0.64, 0.96] \subseteq [0.6, 1.0]$.

For the other tuples, one has $prob_{R,r,t_i}(x.\text{P\_DISEASE} = cirrhosis) = [0, 0] \not\subset [0.4, 0.7]$, $\forall i \neq 2$. Thus, those tuples do not satisfy ϕ.

4.2 Projection

A projection of a UFRDB relation on a set of attributes is a new UFRDB relation where only the attributes in that set are considered for each tuple of the new relation. Moreover, equivalent tuples under a chosen threshold should be coalesced into a tuple in the result relation by probabilistic combination strategies. The projection operation of a UFRDB relation is extended from the projection operation of a CRDB relation with uncertain and imprecise values of relational tuples and is defined as follows.

Definition 18. Let $R(\boldsymbol{U}, \wp)$ be a UFRDB schema, r be a relation over R, $\boldsymbol{L}$ be a subset of attributes of $\boldsymbol{U}$, $\oplus$ and $\otimes$ be probabilistic disjunction and conjunction strategies with respect to the same combination alternative, $\varepsilon \in [0, 1]$ be an equivalent threshold on $\boldsymbol{L}$. The *projection* of r on $\boldsymbol{L}$ under $\oplus$, $\otimes$ and ε, denoted by $\Pi_{\boldsymbol{L}\oplus\varepsilon\otimes}(r)$, is the relation r^* over the schema R^* determined by:

1. $R^* = (\boldsymbol{L}, \wp^*)$ and $\wp^*(A) = \wp(A)$, $\forall A \in \boldsymbol{L}$.
2. $r^* = \{t^* \mid t^*.A = u.A \oplus \ldots \oplus w.A, \forall A \in \boldsymbol{L}, \exists\, u, \ldots, w \in r$ such that $\forall\, t_i, t_j \in \{u, \ldots, w\}, t_i[\boldsymbol{L}] \approx_{\varepsilon\otimes} t_j[\boldsymbol{L}]\}$.

We note that the combination alternative of a probabilistic combination strategy can be the "ignorance", "independence", "positive correlation" or "mutual exclusion" as in Table 1.

Example 10. Consider the relation DIAGNOSE over the schema **DIAGNOSE**$(\boldsymbol{U}, \wp)$ as in Table 3, where $\boldsymbol{U}$ = {P_ID, D_ ID, P_AGE, P_DISEASE} and *middle_aged*, *approx_40* are the fuzzy sets given in Examples 4. The set $\wp(A)$ for each attribute A in the schema **DIAGNOSE**$(\boldsymbol{U}, \wp)$ consists of all fuzzy probabilistic triples $\langle V, \alpha, \beta\rangle$ on *dom*(A). Then the projection of DIAGNOSE on $\boldsymbol{L}$ = {D_ID, P_AGE, P_DISEASE} under $\oplus_{in}$, $\otimes_{in}$ and the equivalent threshold $\varepsilon = 0.2$ is the relation $r^* = \Pi_{\boldsymbol{L}\oplus_{in}0.2\otimes_{in}}$(DIAGNOSE) over the schema $R^* = (\boldsymbol{L}, \wp^*)$ computed as in Table 4, where $\wp^*(A) = \wp(A)$, $\forall A \in \boldsymbol{L}$.

Table 3: Relation DIAGNOSE

P_ID	D_ID	P_AGE	P_DISEASE
P388	*D102*	⟨{30}, *u*, *u*⟩	⟨{*hepatitis*, *gall-stone*}, 0.8*u*, 1.2*u*⟩
P245	*D025*	⟨{*middle_aged*, *approx_40*}, 0.9*u*, 1.4*u*⟩	⟨{*cholecystitis*}, *u*, *u*⟩
P237	*D102*	⟨{30, 31}, *u*, *u*⟩	⟨{*hepatitis*}, *u*, *u*⟩

Table 4: Relation $\Pi_{L\oplus_{in}0.2\otimes_{in}}$(DIAGNOSE)

D_ID	P_AGE	P_DISEASE
D102	⟨{30, 31}, α, β⟩, where $\alpha(30) = \beta(30)=1$, $\alpha(31)= \beta(31)=0.5$	⟨{*hepatitis*, *gall-stone*}, α, β⟩, where α(*hepatitis*) = β(*hepatitis*)=1, α(*gall-stone*) = 0.4, β(*gall-stone*)=0.6
D025	⟨{*middle_aged*, *approx_40*}, 0.9*u*, 1.4*u*⟩	⟨{*cholecystitis*}, *u*, *u*⟩

We note that two tuples t_1 and t_3 in Table 3 are equivalent on $\boldsymbol{L}$ = {D_ID, P_AGE, P_DISEASE} under the threshold $\varepsilon = 0.2$ and the independence probabilistic conjunction strategy $\otimes_{in}$ and they are projected on $\boldsymbol{L}$ and coalesced into the tuple t_1 under the independence probabilistic disjunction strategy $\oplus_{in}$ in Table 4. However, if we chose another equivalent threshold $\varepsilon > 0.3$, for instance $\varepsilon = 0.5$, then there does not exist any equivalent tuples on $\boldsymbol{L}$ under ε and the result of the projection operation is the relation $\Pi_{\boldsymbol{L}\oplus_{in}0.5\otimes_{in}}$(DIAGNOSE) as in Table 5.

Table 5: Relation $\Pi_{L\oplus_{in}0.5\otimes_{in}}$(DIAGNOSE)

D_ID	P_AGE	P_DISEASE
D102	⟨{30}, *u*, *u*⟩	⟨{*hepatitis*, *gall-stone*}, 0.8*u*, 1.2*u*⟩
D025	⟨{*middle_aged*, *approx_40*}, 0.9*u*, 1.4*u*⟩	⟨{*cholecystitis*}, *u*, *u*⟩
D102	⟨{30, 31}, *u*, *u*⟩	⟨{*hepatitis*}, *u*, *u*⟩

4.3 Cartesian Product

For the Cartesian product of two UFRDB relations, as in CRDB, we assume the set of attributes of their schemas are disjoint and every k-tuple $t = (\langle V_1, \alpha_1, \beta_1\rangle, \ldots, \langle V_k, \alpha_k, \beta_k\rangle)$ is an un-ordered list. The Cartesian product of two UFRDB relations is extended from the Cartesian product of two CRDB relations as follows.

Definition 19. Let $\boldsymbol{U_1}$, $\boldsymbol{U_2}$ be two sets of attributes that have not any common element, $R_1(\boldsymbol{U_1}, \wp_1)$, $R_2(\boldsymbol{U_2}, \wp_2)$ be two UFRDB schemas, r_1, r_2 be two relations over R_1 and R_2, respectively. The *Cartesian product* of r_1 and r_2, denoted by $r_1 \times r_2$, is the relation r over R, determined by:

1. $R = (\boldsymbol{U}, \wp)$, where $\boldsymbol{U} = \boldsymbol{U_1} \cup \boldsymbol{U_2}$, $\wp(A) = \wp_1(A)$ if $A \in \boldsymbol{U}_1$ and $\wp(A) = \wp_2(A)$ if $A \in \boldsymbol{U}_2$.
2. $r = \{t \mid t.A = t_1.A$ if $A \in \boldsymbol{U_1}$, $t.A = t_2.A$ if $A \in \boldsymbol{U_2}$, $t_1 \in r_1$, $t_2 \in r_2\}$.

4.4 Join

The join of two UFRDB relations is extended from the natural join of two CRDB relations with probability and fuzzy set values as following definition.

Definition 20. Let $\boldsymbol{U_1}$ and $\boldsymbol{U_2}$ be two sets of attributes such that if they have the same name attributes, respectively in those two sets then such attributes have the same value domain. Let $R_1(\boldsymbol{U_1}, \wp_1)$ and $R_2(\boldsymbol{U_2}, \wp_2)$ be two UFRDB schemas, r_1, r_2 be two relations over R_1 and R_2, respectively and $\otimes$ be a probabilistic conjunction strategy. The *join* of r_1 and r_2 under $\otimes$, denoted by $r_1 \bowtie_\otimes r_2$, is the relation r over the schema R, determined by:

1. $R = (\boldsymbol{U}, \wp)$ where $\boldsymbol{U} = \boldsymbol{U_1} \cup \boldsymbol{U_2}$, $\wp(A) = \wp_1(A)$ if $A \in \boldsymbol{U_1} - \boldsymbol{U_2}$, $\wp(A) = \wp_2(A)$ if $A \in \boldsymbol{U_2} - \boldsymbol{U_1}$ and $\wp(A) = \wp_1(A) = \wp_2(A)$ if $A \in \boldsymbol{U_1} \cap \boldsymbol{U_2}$.
2. $r = \{t \mid t.A = t_1.A$ if $A \in \boldsymbol{U_1} - \boldsymbol{U_2}$, $t.A = t_2.A$ if $A \in \boldsymbol{U_2} - \boldsymbol{U_1}$, $t.A = t_1.A \otimes t_2.A$ if $A \in \boldsymbol{U_1} \cap \boldsymbol{U_2}$ and $t_1.A \otimes t_2.A \neq \langle \emptyset, \alpha, \beta \rangle$, $t_1 \in r_1$, $t_2 \in r_2\}$.

Example 11. Given two UFRDB relations DOCTOR$_1$ and DOCTOR$_2$ as in Tables 6 and 7, where *young*, *approx_40* and *middle_aged* are the fuzzy sets given in Example 4. Then, the result of the join of them under the probabilistic conjunction strategy $\otimes_{in}$ and the *standard intersection* of fuzzy sets (by Definition 5) is the relation DOCTOR$_1$ $\bowtie_{\otimes in}$ DOCTOR$_2$ computed as in Table 8. Here, the names of each relation and its schema are identical, the set of fuzzy probabilistic triples $\wp(A)$ for each attribute A in the schemas consists of all fuzzy probabilistic triples on *dom*(A).

Table 6: Relation DOCTOR$_1$

D_ID	D_AGE
D005	$\langle$\{*middle_aged*, *approx_40*\}, 0.7*u*, 1.3*u*$\rangle$
D093	$\langle$\{*young*\}, *u*, *u*$\rangle$
D102	$\langle$\{55, 56\}, *u*, *u*$\rangle$

Table 7: Relation DOCTOR$_2$

D_NAME	D_AGE
Alice	$\langle$\{30, 31\}, 0.8*u*, 1.2*u*$\rangle$
George	$\langle$\{*approx_40*\}, *u*, *u*$\rangle$
Peter	$\langle$\{54, 55\}, *u*, *u*$\rangle$

Table 8: Relation DOCTOR$_1$ $\bowtie_{\otimes in}$ DOCTOR$_2$

D_ID	D_NAME	D_AGE
D005	George	$\langle$\{*approx_40*\}, 0.35*u*, 0.65*u*$\rangle$
D102	Peter	$\langle$\{55\}, 0.25*u*, 0.25*u*$\rangle$

We note that *middle_aged* $\cap$ *approx_40* = *approx_40*, so $\langle$\{*middle_aged*, *approx_40*\}, 0.7*u*, 1.3*u*$\rangle$ $\otimes_{in}$ $\langle$\{*approx_40*\}, *u*, *u*$\rangle$ = $\langle$\{*approx_40*\}, 0.35*u*, 0.65*u*$\rangle$. Consequently, the tuple t_1 in Table 8 is the result of the join of the tuple t_1 in Table 6 and the tuple t_2 in Table 7.

4.5 Intersection, Union and Difference

The intersection, union and difference of two UFRDB relations over the same schema is a UFRDB relation over that schema, where two equivalent tuples under a threshold ε, respectively of those two relations are coalesced into a tuple in the result relation by a probabilistic combination strategy. Thus, the operations are an extension of the intersection, union and difference of two CRDB relations with probability and fuzzy set values. The intersection, union and difference of two UFRDB relations in turn are defined as below.

Definition 21. Let $R(\boldsymbol{U}, \wp)$ be a UFRDB schema, r_1 and r_2 be two relations over R, $\otimes$ be a probabilistic conjunction strategy, and $\varepsilon \in [0, 1]$ be an equivalent threshold on $\boldsymbol{U}$. The *intersection* of r_1 and r_2 under $\otimes$ and ε, denoted by $r_1 \cap_{\varepsilon\otimes} r_2$, is the relation r over $R(\boldsymbol{U}, \wp)$ defined by $r = \{t \mid t.A = t_1.A \otimes t_2.A$, $t_1 \in r_1$, $t_2 \in r_2$, $A \in \boldsymbol{U}$, such that $t_1 \approx_{\varepsilon\otimes} t_2$ and $t_1.A \otimes t_2.A \neq \langle \emptyset, \alpha, \beta \rangle\}$.

Example 12. Consider two UFRDB relations DIAGNOSE$_1$ and DIAGNOSE$_2$ over the same schema **DIAGNOSE**($\boldsymbol{U}$, $\wp$) as in Tables 9 and 10, where $\boldsymbol{U}$ = \{P_ID, D_ID, P_AGE, P_DISEASE\}, *approx_42* = \{41: 0.9, 42: 1.0, 43: 0.9\}, *young*, *approx_40* and *middle_aged* are fuzzy set given in Example 4. The set $\wp(A)$ for each attribute A in the schema **DIAGNOSE**($\boldsymbol{U}$, $\wp$) consists of all fuzzy probabilistic triples $\langle V, \alpha, \beta \rangle$ on *dom*(A). Then the intersection of DIAGNOSE$_1$ and DIAGNOSE$_2$ under $\otimes_{in}$ and the equivalent threshold ε = 0.194 is the relation DIAGNOSE$_1$$\cap_{0.194\otimes in}$DIAGNOSE$_2$ computed as in Table 11.

Here, we note that the tuple t_1 in Table 9 and the tuple t_2 in Table 10 are equivalent on $\boldsymbol{U}$ = \{P_ID, D_ID, P_AGE, P_DISEASE\} under the threshold ε = 0.194 and the independence probabilistic conjunction strategy $\otimes_{in}$ (because $\otimes_{in\, A \in \boldsymbol{U}}$ $p(t_1.A =_{\otimes in} t_2.A) = [1, 1] \otimes_{in} [1, 1] \otimes_{in} [0.194, 0.301] \otimes_{in} [1, 1] = [0.194, 0.301]$), consequently they are coalesced into the tuple

t_1 under $\otimes_{in}$ in the Table 11. In addition, it easy to see that the tuple t_2 in Table 11 is the result of coalescence of the tuple t_3 in Table 9 and the tuple t_1 in Table 10.

Table 9: Relation DIAGNOSE$_1$

P_ID	D_ ID	P_AGE	P_DISEASE
P234	*D102*	⟨{*approx_40*}, *u*, *u*⟩	⟨{*hepatitis*}, *u*, *u*⟩
P217	*D093*	⟨{*middle_aged*, *approx_40*}, 0.6*u*, 1.2*u*⟩	⟨{*lung cancer*, *tuberculosis*}, 0.8*u*, 1.2*u*⟩
P383	*D105*	⟨{69}, *u*, *u*⟩	⟨{*lung cancer*}, *u*, *u*⟩

Table 10: Relation DIAGNOSE$_2$

P_ID	D_ ID	P_AGE	P_DISEASE
P383	*D105*	⟨{69}, *u*, *u*⟩	⟨{*lung cancer*}, *u*, *u*⟩
P234	*D102*	⟨{*approx_40*, *approx_42*}, 0.9*u*, 1.4*u*⟩	⟨{*hepatitis*}, *u*, *u*⟩
P242	*D025*	⟨{*young*}, *u*, *u*⟩	⟨{*cholecystitis*, *cirrhosis*}, 0.7*u*, 1.3*u*⟩

Table 11: Relation DIAGNOSE$_1\cap_{0.194\otimes in}$DIAGNOSE$_2$

P_ID	D_ ID	P_AGE	P_DISEASE
P234	*D102*	⟨{*approx_40*}, 0.45*u*, 0.7*u*⟩	⟨{*hepatitis*}, *u*, *u*⟩
P383	*D105*	⟨{69}, *u*, *u*⟩	⟨{*lung cancer*}, *u*, *u*⟩

However, if we chose the equivalent threshold ε =1.0, then only the tuple t_3 in Table 9 and the tuple t_1 in Table 10 are equivalent on U under ε and the result of the intersection operation is the relation DIAGNOSE$_1\cap_{1.0\otimes in}$DIAGNOSE$_2$ as in Table 12.

Table 12: Relation DIAGNOSE$_1\cap_{1.0\otimes in}$DIAGNOSE$_2$

P_ID	D_ ID	P_AGE	P_DISEASE
P383	*D105*	⟨{69}, *u*, *u*⟩	⟨{*lung cancer*}, *u*, *u*⟩

Definition 22. Let $R(\boldsymbol{U}, \wp)$ be a UFRDB schema, r_1 and r_2 be two relations over R, $\oplus$ and $\otimes$ be probabilistic disjunction and conjunction strategies with respect to the same combination alternative, and $\varepsilon\in[0, 1]$ be an equivalent threshold on $\boldsymbol{U}$. The *union* of r_1 and r_2 under $\oplus$, $\otimes$ and ε, denoted by $r_1\cup_{\varepsilon\oplus\otimes} r_2$, is the relation r over $R(\boldsymbol{U}, \wp)$ defined by $r = \{t_1 \in r_1 \mid$ there is not any tuple $t_2 \in r_2$ such that $t_1 \approx_{\varepsilon\otimes} t_2\}\cup\{t_2 \in r_2 \mid$ there is not any tuple $t_1 \in r_1$ such that $t_2 \approx_{\varepsilon\otimes} t_1\}\cup\{t \mid t.A = t_1.A \oplus t_2.A, t_1 \in r_1, t_2 \in r_2, A \in \boldsymbol{U}$ such that $t_1 \approx_{\varepsilon\otimes} t_2\}$.

Definition 23. Let $R(\boldsymbol{U}, \wp)$ be a UFRDB schema, r_1 and r_2 be two relations over R, $\ominus$ and $\otimes$ be probabilistic difference and conjunction strategies with respect to the same combination alternative, and $\varepsilon \in [0, 1]$ be an equivalent threshold on $\boldsymbol{U}$. The *difference* of r_1 and r_2 under $\ominus$, $\otimes$ and ε, denoted by $r_1 -_{\varepsilon\ominus\otimes} r_2$, is the relation r over $R(\boldsymbol{U}, \wp)$ defined by $r = \{t_1 \in r_1 \mid$ there is not any tuple $t_2 \in r_2$ such that $t_1 \approx_{\varepsilon\otimes} t_2\}\cup\{t \mid t.A = t_1.A \ominus t_2.A, t_1 \in r_1, t_2 \in r_2, A \in \boldsymbol{U}$ such that $t_1 \approx_{\varepsilon\otimes} t_2$ and $t_1.A \ominus t_2.A \neq \langle\varnothing, \alpha, \beta\rangle\}$.

Example 13. Given two UFRDB relations DIAGNOSE$_1$ and DIAGNOSE$_2$ over the same schema **DIAGNOSE**$(\boldsymbol{U}, \wp)$ as in Tables 9 and 10 of Example 12. Then the difference of DIAGNOSE$_1$ and DIAGNOSE$_2$ under $\ominus_{in}$, $\otimes_{in}$ and the equivalent threshold ε = 0.194 is the relation DIAGNOSE$_1-_{0.194\ominus in\otimes in}$DIAGNOSE$_2$ computed as in Table 13.

Table 13: Relation DIAGNOSE$_1-_{0.194\ominus in\otimes in}$DIAGNOSE$_2$

P_ID	D_ ID	P_AGE	P_DISEASE
P217	*D093*	⟨{*middle_aged*, *approx_40*}, 0.6*u*, 1.2*u*⟩	⟨{*lung cancer*, *tuberculosis*}, 0.8*u*, 1.2*u*⟩

Meanwhile, if the chosen equivalent threshold ε > 0.194, for instance ε = 1.0, then the result relation DIAGNOSE$_1-_{1.0\ominus in\otimes in}$DIAGNOSE$_2$ computed as in Table 14.

Table 14: Relation DIAGNOSE$_1-_{1.0\ominus in\otimes in}$DIAGNOSE$_2$

P_ID	D_ ID	P_AGE	P_DISEASE
P234	*D102*	⟨{*approx_40*}, *u*, *u*⟩	⟨{*hepatitis*}, *u*, *u*⟩
P217	*D093*	⟨{*middle_aged*, *approx_40*}, 0.6*u*, 1.2*u*⟩	⟨{*lung cancer*, *tuberculosis*}, 0.8*u*, 1.2*u*⟩

5. PROPERTY OF ALGEBRAIC OPERATIONS

In this section, we propose some properties of the UFRDB algebraic operations as an extension from those in CRDB. Clearly, these properties say that our UFRDB model is coherent and consistent.

Proposition 1. Let *R* be a UFRDB schema, *r* be a relation over *R*, ϕ_1 and ϕ_2 be two selection conditions. Then

$$\sigma_{\phi_1}(\sigma_{\phi_2}(r)) = \sigma_{\phi_2}(\sigma_{\phi_1}(r)) = \sigma_{\phi_1\wedge\phi_2}(r) \quad (1)$$

where, the last expression assumes that ϕ_1 and ϕ_2 have the same tuple variable.

Proof: Let $r_1 = \sigma_{\phi_1}(r)$, $r_2 = \sigma_{\phi_2}(r)$ and $r_{1\wedge 2} = \sigma_{\phi_1\wedge\phi_2}(r)$. Then for each $t \in r$, we have

$$\begin{aligned}\sigma_{\phi_1}(\sigma_{\phi_2}(r)) &= \{t \in r_2 \mid prob_{R,r_2,t} \vDash \phi_1\}\\ &= \{t \in r \mid (prob_{R,r,t} \vDash \phi_2) \wedge (prob_{R,r_2,t} \vDash \phi_1)\}\\ &= \{t \in r \mid (prob_{R,r,t} \vDash \phi_2) \wedge (prob_{R,r,t} \vDash \phi_1)\} \quad \text{(because of } r_2 \subseteq r)\\ &= \{t \in r \mid prob_{R,r,t} \vDash \phi_1 \wedge \phi_2\} \quad \text{(Definition 16)}\\ &= \sigma_{\phi_1\wedge\phi_2}(r).\end{aligned}$$

So, $\sigma_{\phi_1}(\sigma_{\phi_2}(r)) = \sigma_{\phi_1\wedge\phi_2}(r)$ is proven. The equation $\sigma_{\phi_2}(\sigma_{\phi_1}(r)) = \sigma_{\phi_2\wedge\phi_1}(r)$ is proven similarly. Since $\phi_1\wedge\phi_2 \Leftrightarrow \phi_2\wedge\phi_1$ (the logical conjunction of selection conditions are commutative), hence $\sigma_{\phi_1\wedge\phi_2}(r) = \sigma_{\phi_2\wedge\phi_1}(r)$. Therefore, we have $\sigma_{\phi_1}(\sigma_{\phi_2}(r)) = \sigma_{\phi_2}(\sigma_{\phi_1}(r))$ and so $\sigma_{\phi_1}(\sigma_{\phi_2}(r)) = \sigma_{\phi_2}(\sigma_{\phi_1}(r)) = \sigma_{\phi_1\wedge\phi_2}(r)$. Thus, Proposition 1 is proven.

Proposition 2. Let *R* be a UFRDB schema, *r* be a relation over *R*, $\oplus$ and $\otimes$ be probabilistic disjunction and conjunction strategies with respect to the same combination alternative, ***A*** and ***B*** be two subsets of attributes of *R*, $\boldsymbol{A} \subseteq \boldsymbol{B}$ and $\varepsilon \in [0, 1]$ be an equivalent threshold on ***B***. Then

$$\Pi_{\boldsymbol{A}\oplus\varepsilon\otimes}(\Pi_{\boldsymbol{B}\oplus\varepsilon\otimes}(r)) = \Pi_{\boldsymbol{A}\oplus\varepsilon\otimes}(r) \quad (2)$$

Proof: Because $\boldsymbol{A} \subseteq \boldsymbol{B}$, so $\boldsymbol{A}\cap\boldsymbol{B} = \boldsymbol{A}$ and sides of (2) are the relations over the same schema (Definition 18). Moreover, it is due to $\boldsymbol{A} \subseteq \boldsymbol{B}$, so ε-equivalent tuples on ***B*** are also ε-equivalent on ***A*** with respect to $\otimes$ (Definition 12). From that, we are easy to see $\Pi_{\boldsymbol{A}\oplus\varepsilon\otimes}(\Pi_{\boldsymbol{B}\oplus\varepsilon\otimes}(r)) = \Pi_{\boldsymbol{A}\cap\boldsymbol{B}\oplus\varepsilon\otimes}(r) = \Pi_{\boldsymbol{A}\oplus\varepsilon\otimes}(r)$ under the equivalent threshold ε and the same combination alternative of $\oplus$ and $\otimes$. Thus, the equation (2) is proven.

Proposition 3. Let R_1, R_2 and R_3 be the UFRDB schemas such that if they have the same name attributes then such attributes have the same value domain, r_1, r_2 and r_3 be relations over R_1, R_2 and R_3 respectively, $\otimes$ be a probabilistic conjunction strategy. Then

$$r_1 \bowtie_\otimes r_2 = r_2 \bowtie_\otimes r_1 \quad (3)$$

$$(r_1 \bowtie_\otimes r_2) \bowtie_\otimes r_3 = r_1 \bowtie_\otimes (r_2 \bowtie_\otimes r_3) \quad (4)$$

The equations (3) and (4) say that the join operation of UFRDB relations is commutative and associative.

Proof: Clearly, $r_1 \bowtie_\otimes r_2$ and $r_2 \bowtie_\otimes r_1$ are two relations over the same schema. By Definition 5, the conjunction of fuzzy probabilistic triples is commutative (due to the commutativity of probabilistic conjunction strategies and the intersection of fuzzy sets). So, by Definition 20, we have $r_1 \bowtie_\otimes r_2 = r_2 \bowtie_\otimes r_1$.

By Definition 20, the results of two sides of (4) are the relations over the same schema. Moreover, the intersection of fuzzy sets has the associativity, by Definition 5, it follows that the conjunction of fuzzy probabilistic triples is associative. From the associativity of the classical relational join and by Definition 20, it is easy to see that the join of UFRDB relations is associative. Thus, it results in $(r_1 \bowtie_\otimes r_2) \bowtie_\otimes r_3 = r_1 \bowtie_\otimes (r_2 \bowtie_\otimes r_3)$.

Because the Cartesian product is a particular case of the join (Definition 20), we have the straight corollary of Proposition 3 below.

Corollary 1. Let R_1, R_2 and R_3 be UFRDB schemas such that each pair of them has not any common attribute, r_1, r_2 and r_3 be relations over R_1, R_2 and R_3 respectively. Then

$$r_1 \times r_2 = r_2 \times r_1 \quad (5)$$

$$(r_1 \times r_2) \times r_3 = r_1 \times (r_2 \times r_3) \quad (6)$$

Proposition 4. Let *R* be a UFRDB schema, r_1, r_2 and r_3 be relations over *R*, $\otimes$ and $\oplus$ be probabilistic conjunction and disjunction strategies with respect to the same combination alternative, $\varepsilon \in [0, 1]$. Then

$$r_1 \cap_{\varepsilon\otimes} r_2 = r_2 \cap_{\varepsilon\otimes} r_1 \quad (7)$$

$$(r_1 \cap_{\varepsilon\otimes} r_2) \cap_{\varepsilon\otimes} r_3 = r_1 \cap_{\varepsilon\otimes} (r_2 \cap_{\varepsilon\otimes} r_3) \quad (8)$$

$$r_1 \cup_{\varepsilon\oplus\otimes} r_2 = r_2 \cup_{\varepsilon\oplus\otimes} r_1 \quad (9)$$

$$(r_1 \cup_{\varepsilon\oplus\otimes} r_2) \cup_{\varepsilon\oplus\otimes} r_3 = r_1 \cup_{\varepsilon\oplus\otimes} (r_2 \cup_{\varepsilon\oplus\otimes} r_3) \quad (10)$$

The equations of (7), (8), (9) and (10) say that the intersection and union of relations in UFRDB are commutative and associative.

Proof: The equations in the proposition are proven respectively as follows:

Equations (7) and (8): For every equivalent threshold ε chosen, then the equivalent tuples in relations do not change. Moreover, from the commutativity and associativity of the intersection of fuzzy sets, it follows the commutativity and associativity of the conjunction of fuzzy probabilistic triples and the commutativity and associativity of the intersection of UFRDB relations under the equivalent threshold ε and the probabilistic conjunction strategy ⊗. From that and by Definition 21, it follows Equations (7) and (8).

Equations (9) and (10): As for the equations (7) and (8), under an equivalent threshold ε chosen, then the equivalent tuples in relations do not change. From the commutativity and associativity of the union, intersection of fuzzy sets, the disjunction of fuzzy probabilistic triples (Definition 6), by Definition 22, it follows the union of UFRDB relations being commutative and associative under the equivalent threshold ε and the same combination alternative of ⊕ and ⊗. Thus, we have the equations (9) and (10).

6. CONCLUSION

In this paper, we have proposed a hybrid relational database model, called UFRDB, for representing and manipulating imprecise and uncertain information. UFRDB has been built by extending and generalizing the classical relational database model, where the relational attribute value is defined as a fuzzy probabilistic triple, the computation and combination of relational tuples are implemented by using the mass assignment, the probabilistic interpretation of binary relations on fuzzy sets and the combination strategies of fuzzy probabilistic triples. A notion of the equivalence of relational tuples has been proposed for eliminating redundant tuples and the consistency of relatons. The data model and basic relational algebraic operations for UFRDB have been formally defined accordingly. A set of basic properties of the algebraic operations in UFRDB have also been proposed and proven completely.

Towards applying UFRDB in practice, we will build a management system for UFRDB with the familiar querying and manipulating language like SQL that is able to represent and handle imprecise and uncertain information in the real world.

7. REFERENCES

[1] J.F. Baldwin, J.M. Lawry, and T.P. Martin, "A mass assignment theory of the probability of fuzzy events", Journal of Fuzzy Sets and Systems, vol. 83, pp. 353-367, 1996.
[2] J.F. Baldwin, J.M. Lawry, and T.P. Martin, "A note on probability/possibility consistency for fuzzy events", in Proceedings of the 6th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU'96, Granada, Spain, pp. 521-525, 1996.
[3] E.F. Codd, "A relational model of data for large shared data banks", Communications of the ACM, vol. 13, pp. 377-387, 1970.
[4] T. H. Cao, and H. Nguyen, "Uncertain and fuzzy object bases: a data model and algebraic operations", International Journal of Uncertainty Fuzziness and Knowledge-Based Systems, vol. 19, pp. 275–305, 2011.
[5] C.J. Date, An introduction to database systems, the 8th Edition, Addison Wesley, 2008.
[6] T. Cao, H. Nguyen, A. Inoue, and D. Zhou, "A Probabilistic Relational Database Model with Fuzzy Attribute Values", in Proceedings of the 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, Louisiana,USA, pp.338-343, 2019.
[7] D. Dey and S. A. Sarkar, "A probabilistic relational model and algebra", ACM Transactions on Database Systems, vol. 21, pp. 339-369, 1996.
[8] T. Eiter, T. Lukasiewicz, and M. Walter, "A data model and algebra for probabilistic complex values", Annals of Mathematics and Artificial Intelligence, vol. 33, pp. 205-252, 2001.
[9] T. Eiter, J.J. Lu, T. Lukasiewicz, and V.S. Subrahmanian, "Probabilistic object bases", ACM Transactions on Database Systems (TODS), vol. 26, pp. 264-312, 2001.
[10] T. Ge, A. Dekhtyar, and J. Goldsmith, "Uncertain data: Representations, query processing, and applications", in Studies in Fuzziness and Soft Computing, Springer-Verlag Berlin Heidelberg, pp. 67-108, 2013.
[11] G.J. Klir, and B. Yuan, Fuzzy sets and fuzzy logic -Theory and applications, Prentice Hall PTR, 1994.
[12] Z. Ma, and L. Yan, Advances in probabilistic databases for uncertain information management, Springer, 2013.
[13] Z. Ma, and L. Yan, "Modeling fuzzy data with RDF and fuzzy relational database models", International Journal of Intelligent Systems, vol. 33, pp. 1534-1554, 2018.
[14] X. Meng, Z.M. Ma, and X. Zhu, "A Knowledge-based fuzzy query and results ranking approach for relational databases", Journal of Computational Information Systems, vol. 6, pp. 2037-2044, 2010.
[15] J. Mishra, and S. Ghosh, "A new functional dependency in a vague relational database model", International Journal of Computer Applications, vol. 39, pp. 29-36, 2012.
[16] H. Nguyen, and D.H. Tran, "A probabilistic relational data model for uncertain information", in Proceedings of the 3rd IEEE International Conference on Information Science and Technology (ICIST), Yangzhou, China, pp. 607-613, 2013.
[17] H. Nguyen, "Towards a fuzzy probabilistic relational data base model", in Proceedings of the 7th IEEE International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, Vietnam, pp. 298-301, 2015.
[18] H. Nguyen, "A fuzzy probabilistic relational database model and algebra", Journal of Computer Science and Cybernetics, vol. 32, pp. 189-207, 2016.

[19] H. Nguyen, “Extending relational database model for uncertain information”, Journal of Computer Science and Cybernetics, vol.35, pp.355-372, 2019.
[20] K. Papaioannou, M. Theobald, and M. Boehlen, “Supporting set operations in temporal-probabilistic databases”, in Proceedings of the 34th IEEE International Conference on Data Engineering (ICDE), Paris, France, pp. 1180 -1191, 2018.
[21] F.E. Petry, Fuzzy databases: Principles and applications, Kluwer Academic Publishers, 1996.
[22] R. Ross, and V.S. Subrahmanian, “Aggregate operators in probabilistic databases”, Journal of the ACM, vol. 52, pp. 54-101, 2005.
[23] A.A. Sabour, A.M. Gadallah, and H.A. Hefny, “Flexible querying of relational databases: fuzzy set based approach”, Advanced Machine Learning Technologies and Applications Communications in Computer and Information Science, vol. 488, pp. 446-455, 2014.
[24] L. Yan, and Z.M. Ma, “A fuzzy probabilistic relational database model and algebra”, International Journal of Fuzzy Systems, vol. 15, pp. 244-253, 2013.
[25] L. Yan, and Z.M. Ma, “A Probabilistic Nested Relational Database Model with Fuzzy Probability Measures”, in Proceedings of the 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, Louisiana,USA, pp.253-257, 2019.
[26] L.A. Zadeh, “Fuzzy sets”, Journal of Information and Control, vol. 8, pp. 338-353, 1965.
[27] S. Zhang, and C. Zhang, “A probabilistic data model and its semantics”, Journal of Research and Practice in Information Technology, vol. 35, pp. 237-256, 2003.
[28] W. Zhao, A. Dekhtyar, and J. Goldsmith, “Databases for interval probabilities”, International Journal of Intelligent Systems, vol. 19, pp. 789-815, 2004.