Journal of Computer Science and Cybernetics, V.32, N.2 (2016), 189–207
DOI: 10.15625/1813-9663/32/2/7887

A FUZZY PROBABILISTIC RELATIONAL DATABASE MODEL AND ALGEBRA

NGUYEN HOA

Department of Information Technology, Saigon University; nguyenhoa@sgu.edu.vn

Abstract. This paper introduces a complete extended relational database model based on probability theory and fuzzy theory, called FPRDB, for representing and handling vague and uncertain information of objects in real world applications. FPRDB is built by first extending probabilistic triples with fuzzy sets, associating the probability of such fuzzy sets for expressing and computing uncertain degree of imprecise information. Then, schemas, fuzzy probabilistic relations, fuzzy probabilistic functional dependencies and algebraic operations are defined coherently and consistently for FPRDB. A set of the properties of the relational algebraic operations in FPRDB is also formulated and proven.

Keywords. Probability distribution, fuzzy probabilistic triple, fuzzy probabilistic relation, fuzzy probabilistic functional dependency, fuzzy probabilistic relational algebraic operation.

1. INTRODUCTION

It is true that the real world is pervaded by uncertain and imprecise information that we have to face, and make decisions on, in daily life [1, 2]. Although, the classical relational database model is very useful for modeling, designing and implementing large-scale systems, it is restricted for representing and handling uncertain and imprecise information [3, 4]. For example, applications of the classical relational database model cannot deal with queries such as "find all patients who are *young*"; nor "find all players who are 80-90% likely to be the top scorers of the English Premier League, in year 2016", etc, where *young* is an imprecise notion [5, 6].

Up to now, there have been many relational database models studied and built based on the probability theory for modeling objects about which information may be uncertain to overcome the limitation of the classical database models. Such models [7–11] are called *probabilistic relational database models*. These models did not represent and handle vague information of objects. A large number of another relational database models has been built based on the fuzzy set theory for modeling objects about which information may be vague. These model sare called *fuzzy relational database models* [5,12–15] and they did not represent and handle probabilistic information of objects.

In the real world, information may contain both uncertainty and impreciseness. For example, the query "find all patients who are *young* and at least 90% likely to catch a cirrhosis or hepatitis" relates to both impreciseness and uncertainty of the information. When the information of objects contains both uncertainty and impreciseness, the above models can not be applied. However, relational database models combining both the fuzzy set theory and probability theory for modeling objects about which information contains both uncertainty and impreciseness have rarely been proposed. Such models are called *fuzzy probabilistic relational database models*.

© 2016 Vietnam Academy of Science & Technology

Recently, in [16] the authors have developed a fuzzy probabilistic relational database model to represent and manipulate uncertain and imprecise information of objects in the real world applications. In this model, each attribute of a tuple was assigned to a single value with a probability that was inferred from the possibility distribution of probability values associated with the tuple. 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 that can be vague of a certain set.

More recently, in [17], the author introduced a probabilistic relational database model called PRDB that 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 takes one of values of a candidate set. It means that the model can overcome the shortcoming of the model in [16]. However, the PRDB model could not express and deal with vague information. In [18], the author extended the model in [17] with fuzzy set values to a fuzzy probabilistic relational data base model, called FPRDB, for representing and handling uncertain and vague information. However, in the model [18], only the selection operation was defined while all other algebraic operations are missing. In this paper, using the combination of fuzzy probabilistic triples introduced in [2] and the probabilistic interpretation of relations on fuzzy sets defined in [18], we extend more the FPRDB model with a complete set of fuzzy probabilistic relational algebraic operations for representing and manipulating both imprecise and uncertain information in practice.

The basis of mathematics to develop FPRDB is presented in Section 2. Schemas and relations of FPRDB are introduced in Section 3. Section 4,5 and 6 present fuzzy probabilistic relational algebraic operations and their properties in FPRDB. Finally, Section 7 concludes the paper and outlines further research directions in the future.

2. PROBABILITY AND FUZZY SETS

In this section, some notions about probability and fuzzy sets are presented as the basis for extending the probabilistic relational database model PRDB with fuzzy set values.

2.1. Probabilistic interpretation of relations on fuzzy sets

First, the mass assignment formulated based on the voting model of fuzzy sets in [19] as the basis for the probabilistic interpretation of relations on them 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 $i < j$, $\forall i = 1,\ldots, n$ and $\forall j = 1,\ldots, m_i$. The *mass assignment corresponding to* A is a mapping m_A: $2^U \to [0, 1]$ that is defined by $m_A(z_1) = y_1 - y_2, \ldots, m_A(\cup_{i=1,j} z_i) = y_j - y_{j+1},\ldots, m_A(\cup_{i=1,n} z_i) = y_n$, where $z_i = \cup_{j=1,m_i}\{x_{i,j}\}$.

It is noted that, the mass assignment

$$m_A(z_1) = y_1 - y_2, \ldots, m_A(\cup_{i=1,j} z_i) = y_j - y_{j+1},\ldots, m_A(\cup_{i=1,n} z_i) = y_n$$

can be denoted by

$$m_A = z_1\text{: } y_1 - y_2, \ldots, \cup_{i=1,j} z_i\text{: } y_j - y_{j+1},\ldots, \cup_{i=1,n} z_i\text{: } y_n.$$

The probabilistic interpretation of relations on fuzzy sets, as the probability measures for the relations are true, is defined in [2,18] 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 $\{=, \neq, \leq, <, \subseteq, \in\}$ assumed to be valid on $(U \times V)$. The *probabilistic interpretation* of a 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 and $p(u\theta v|u \in S,\ v \in T)$ is the conditional probability of $u\theta v$ given $u \in S$, $v \in T$.

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 $A \to B$, denoted by $prob(A \to B)$, 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 and $p(u \in T|u \in S)$ is the conditional probability for $u \in T$ given $u \in S$.

The intuitive meaning of $prob(A \to B)$ is that given a fuzzy proposition "$x \in A$", $prob(A \to B)$ is the probability for $x \in B$ being true. In other words, it is the fuzzy conditional probability of $x \in B$ given $x \in A$. It is noted that the above probabilistic interpretation can also be adapted for fuzzy sets on continuous domains, using integration instead of additionas for computing the probability of fuzzy events in [20].

Example 1. Let $A = \{3{:}0.2,\ 4{:}0.5,\ 5{:}0.9,\ 6{:}1\}$ and $B = \{6{:}\ 0.3,\ 5{:}1,\ 4{:}0.3\}$ be the fuzzy sets on the domain $\{1, 2, 3, 4, 5, 6\}$, then the mass assignments corresponding to A and B are m_A=$\{6\}$:0.1, $\{5, 6\}$:0.4, $\{4, 5, 6\}$:0.3, $\{3, 4, 5, 6\}$:0.2 and $m_B = \{5\}$:0.7, $\{4, 5, 6\}$:0.3. The probabilistic interpretation of $A \to B$ is computed as follows:

$$\begin{aligned} prob(A \to B) = {} & p(u \in\{5\}|u \in\{6\}).m_A(\{6\}).m_B(\{5\}) + \\ & p(u \in\{5\}|u \in\{5{,}6\}).m_A(\{5{,}6\}).m_B(\{5\}) + \\ & p(u \in\{5\}|u \in\{4{,}5{,}6\}).m_A(\{4{,}5{,}6\}).m_B(\{5\}) + \\ & p(u \in\{5\}|u \in\{3{,}4{,}5{,}6\}).m_A(\{3{,}4{,}5{,}6\}).m_B(\{5\}) + \\ & p(u \in\{4{,}5{,}6\}|u \in\{6\}).m_A(\{6\}).m_B(\{4{,}5{,}6\}) + \\ & p(u \in\{4{,}5{,}6\}|u \in\{5{,}6\}).m_A(\{5{,}6\}).m_B(\{4{,}5{,}6\}) + \\ & p(u \in\{4{,}5{,}6\}|u \in\{4{,}5{,}6\}).m_A(\{4{,}5{,}6\}).m_B(\{4{,}5{,}6\}) + \\ & p(u \in\{4{,}5{,}6\}|u \in\{3{,}4{,}5{,}6\}).m_A(\{3{,}4{,}5{,}6\}).m_B(\{4{,}5{,}6\}) \\ = {} & 0 \times 0.1\times 0.7 + 1/2 \times 0.4 \times 0.7 + 1/3 \times 0.3 \times 0.7 + 1/4 \times 0.2 \times 0.7 + \\ & 1.0 \times 0.1 \times 0.3 + 1.0 \times 0.4 \times 0.3 + 1.0 \times 0.3 \times 0.3 + 3/4 \times 0.2 \times 0.3 = 0.53. \end{aligned}$$

2.2. Fuzzy probabilistic triples

For representing imprecise and uncertain attribute values in FPRDB, the fuzzy probabilistic triples are used in [18] extended from probabilistic triples in [17]. First, the probability distribution function as the basis for the concept of the fuzzy probabilistic triple is defined as below.

Definition 4. Let X be a finite set, a *probability distribution function* α over X is a mapping α: $X \to [0, 1]$ such that $\sum_{x\in X} \alpha(x) \leq 1$.

An important probability distribution function which we often encounter in practice is the uniform distribution $u(x) = 1/|X|$, $\forall x \in X$. For example, if $X = \{e_1, e_2, e_3\}$, the uniform distribution u over X is $u(x) = 1/3$, $\forall x \in\{e_1, e_2, e_3\}$.

Definition 5. A *probabilistic triple* $\langle X, \alpha, \beta \rangle$ consists of a finite set X, a probability distribution function α over X, and a function β: $X \to [0, 1]$ such that $\alpha(x) \leq \beta(x)$, $\forall x \in X$. If the elements of X are fuzzy sets then $\langle X, \alpha, \beta \rangle$ is called a *fuzzy probabilistic triple*.

Informally, a fuzzy probabilistic triple $\langle X, \alpha, \beta \rangle$ assigns each element $x \in X$ a probability interval $[\alpha(x), \beta(x)]$ to express the imprecise and uncertainty degree of x in X. It means that the fuzzy probabilistic triple $\langle X, \alpha, \beta \rangle$ assigns each element $x \in X$ a probability $p(x)$ which $\alpha(x) \leq p(x) \leq \beta(x)$ to represent the imprecise and uncertainty degree of x in X. It is easy to see that the fuzzy probabilistic triple $\langle X, \alpha, \beta \rangle$ is a probability interval extension of the probability distribution function α on X. Many probabilistic database models, such as [1, 9, 11, 17], used the interval $[\alpha(x), \beta(x)]$ to represent the probability for x instead of using the value of the distribution function $\alpha(x)$. Because, in many situations, it is impossible to know exactly the probability α for x but only the probability for x belonging to an interval $[\alpha, \beta]$.

Example 2. Suppose the daily treatment cost of a patient is estimated within about 50 or 60 (thousand VND) with a probability for each between 0.4 and 0.6. Then this information can be represented by the fuzzy probabilistic triple $\langle X, \alpha, \beta \rangle = \langle \{about\_50, about\_60\}, 0.8u, 1.2u \rangle$, where *about_50* and *about_60* are fuzzy sets defining the imprecise treatment costs and u is the uniform distribution over $X = \{about\_50, about\_60\}$. Here, $0.8u$ and $1.2u$ are the probability distribution functions α and β respectively with $\alpha(x) = 0.8u(x) = 0.8(1/2) = 0.4$ and $\beta(x) = 1.2u(x) = 1.2(1/2) = 0.6$, $\forall x \in X = \{about\_50, about\_60\}$. In addition, it is noted more that

$$\sum_{x \in X} \alpha(x) = 0.8u(about\_50) + 0.8u(about\_60) = 0.8(1/2) + 0.8(1/2) = 0.4 + 0.4 = 0.8 \leq 1$$

and

$$\alpha(x) = 0.8u(x) = 0.8(1/2) = 0.4 \leq 0.6 = 1.2(1/2) = 1.2u(x) = \beta(x), \forall x \in X.$$

It means that $\alpha(x) \leq \beta(x)$ and $\beta(x) \in [0, 1]$, $\forall x \in X$.

2.3. Probabilistic combination strategies

Let two events e_1 and e_2 have probabilities in the intervals $[L_1, U_1]$ and $[L_2, U_2]$, then probability intervals of the conjunction event $e_1 \wedge e_2$, disjunction event $e_1 \vee e_2$, or 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 [1, 17] as presented in Table 1, where $\otimes$, $\oplus$, and $\ominus$ denote the conjunction, disjunction, and difference operators, respectively.

In following sections, the notation $[L_1, U_1] \leq [L_2, U_2]$ is used to replace $L_1 \leq L_2$ and $U_1 \leq U_2$ whereas the notation $[L_1, U_1] \subseteq [L_2, U_2]$ is used to replace $L_2 \leq L_1$ and $U_1 \leq U_2$.

2.4. Conjunction, disjunction and difference of fuzzy probabilistic triples

For building algebraic operations such as the join, intersection, union and difference of fuzzy probabilistic relations in FPRDB, combination strategies of probabilistic triples in [17] are extended with fuzzy sets. It is noted 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 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 $\otimes$ be a probabilistic conjunction strategy. Then the *conjunction* of fpt_1 and fpt_2 under $\otimes$, denoted by $fpt_1 \otimes fpt_2$, is the fuzzy probabilistic triple $fpt = \langle V, \alpha, \beta \rangle$, such that:

Table 1. Examples of probabilistic combination strategies

Strategy	Operators
Ignorance	$([L_1, U_1] \otimes_{ig} [L_2, U_2]) \equiv [\max(0, L_1 + L_2 - 1), \min(U_1, U_2)]$ $([L_1, U_1] \oplus_{ig} [L_2, U_2]) \equiv [\max(L_1, L_2), \min(1, U_1 + U_2)]$ $([L_1, U_1] \ominus_{ig} [L_2, U_2]) \equiv [\max(0, L_1 - U_2), \min(U_1, 1 - L_2)]$
Independence	$([L_1, U_1] \otimes_{in} [L_2, U_2]) \equiv [L_1 \,.\, L_2, U_1 .\, U_2]$ $([L_1, U_1] \oplus_{in} [L_2, U_2]) \equiv [L_1 + L_2 - (L_1 \,.\, L_2), U_1 + U_2 - (U_1 \,.\, U_2)]$ $([L_1, U_1] \ominus_{in} [L_2, U_2]) \equiv [L_1 .\, (1 - U_2), U_1 .\, (1 - L_2)]$
Positive correlation (when e_1 implies e_2, or e_2 implies e_1)	$([L_1, U_1] \otimes_{pc} [L_2, U_2]) \equiv [\min(L_1, L_2), \min(U_1, U_2)]$ $([L_1, U_1] \oplus_{pc} [L_2, U_2]) \equiv [\max(L_1, L_2), \max(U_1, U_2)]$ $([L_1, U_1] \ominus_{pc} [L_2, U_2]) \equiv [\max(0, L_1 - U_2), \max(0, U_1 - L_2)]$
Mutual exclusion (when e_1 and e_2 are mutually exclusive)	$([L_1, U_1] \otimes_{me} [L_2, U_2]) \equiv [0, 0]$ $([L_1, U_1] \oplus_{me} [L_2, U_2]) \equiv [\min(1, L_1 + L_2), \min(1, U_1 + U_2)]$ $([L_1, U_1] \ominus_{me} [L_2, U_2]) \equiv [L_1, \min(U_1, 1 - L_2)]$

1) $V = \{v = v_1 \cap v_2 | v_1 \in V_1, v_2 \in V_2, h(v) = 1, [\alpha_1(v_1), \beta_1(v_1)] \otimes [\alpha_2(v_2), \beta_2(v_2)] \neq [0,0]\}$, and

2) $[\alpha(v), \beta(v)] = \oplus_{me: v1 \in V1, v2 \in V2, v = v1 \cap v2} [\alpha_1(v_1), \beta_1(v_1)] \otimes [\alpha_2(v_2), \beta_2(v_2)]$, for every $v \in V$, where $\oplus_{me}$ is the mutual exclusion probabilistic disjunction strategy.

It is noted that, unlike the PRDB, in that 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 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 $\oplus$ 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 $P = \{v_1 \in V_1 | \neg \exists v_2 \in V_2 : h(v_1 \cap v_2) = 1\}$, $Q = \{v_2 \in V_2 | \neg \exists v_1 \in V_1 : h(v_1 \cap v_2) = 1\}$, and $R = \{v_1 \cap v_2 | v_1 \in V_1, v_2 \in V_2, h(v_1 \cap v_2) = 1\}$, 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 8. 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 $P = \{v_1 \in V_1 | \neg\exists v_2 \in V_2 : h(v_1 \cap v_2) = 1\}$, $Q = \{v = v_1 \cap v_2 | v_1 \in V_1, v_2 \in V_2, h(v_1 \cap v_2) = 1 \text{ and } [\alpha_1(v_1), \beta_1(v_1)] \ominus [\alpha_2(v_2), \beta_2(v_2)] \neq [0, 0]\}$, 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. SCHEMAS AND FUZZY PROBABILISTIC RELATIONS

3.1. Fuzzy probabilistic relational schemas

Fuzzy probabilistic relational schemas are extended from probabilistic relational schemas in [17]. A fuzzy probabilistic relational schema describes a set of attributes of a set of certain objects of which each attribute is associated with fuzzy probabilistic triples as the following definition.

Definition 9. A *fuzzy probabilistic relational schema* is a pair $R = (\boldsymbol{U}, \wp)$, where

1) $\boldsymbol{U} = \{A_1, A_2, \ldots, A_k\}$ 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 (i.e. each element of $\wp(A)$ is a fuzzy probabilistic triple that has the form $\langle V, \alpha, \beta \rangle$, where V is a subset of the set of all fuzzy sets on the value domain of A).

Note that as in the classical relational database, for simplicity, the notations $R(\boldsymbol{U}, \wp)$ and R can be used to replace $R = (\boldsymbol{U}, \wp)$. In addition, the domain of each attribute A is denoted by $dom(A)$.

3.2. Fuzzy probabilistic relations

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

Definition 10. Let $\boldsymbol{U} = \{A_1, A_2, \ldots, A_k\}$ be a set of k pairwise different attributes. *A fuzzy probabilistic relation* r over the fuzzy probabilistic relational schema $R(\boldsymbol{U}, \wp)$, is a finite set $\{t | t = (\langle V_1, \alpha_1, \beta_1 \rangle, \langle V_2, \alpha_2, \beta_2 \rangle, \ldots, \langle V_k, \alpha_k, \beta_k \rangle)\}$ 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 $f_i = \wp(A_i)$ and $V_i \neq \emptyset$, for every $i = 1, 2, \ldots, k$.

Each element t in the relation r over $R(\boldsymbol{U}, \wp)$ is called a *tuple* on $\boldsymbol{U}$. Each 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, the notation $t.A_i$ denotes $\langle V_i, \alpha_i, \beta_i \rangle$ that is $t.A_i = \langle V_i, \alpha_i, \beta_i \rangle$. For each set of attributes $X \subseteq \{A_1, A_2, \ldots, A_k\}$, the notation $t[X]$ is used to denote the rest of t after eliminating the value of attributes not belonging to X.

From Definition 5, it is noted that, each attribute A_i of a tuple t in the relation r over $R(\boldsymbol{U}, \wp)$ only takes one value v in V_i with a probability $p(v) \in [\alpha_i(v), \beta_i(v)]$. As in [1,3,16], the model FPRDB adopts the closed world assumption (CWA). It means, for each tuple t, every value $v \in dom(A_i) - V_i$ has the probability 0. In addition, each precise value $v \in V_i$ is also 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 $x \neq v$. So, each probabilistic relation in [17] can be considered as a particular fuzzy probabilistic relation by Definition 10.

Example 4. A simple fuzzy probabilistic relation PATIENT (over the schema **PATIENT**) about patients at the clinic of a hospital can be organised as Table 2 In the relation, the attributes P_ID, P_NAME, AGE, DISEASE and D_COST describe the information about the identifier, name, age, disease and daily treatment cost of each patient, respectively. In reality, while diagnosing, the disease of each patient is not always determined certainly by the physicians. Similarly, the treatment cost for patients are also not known definitely even as the patients know about their diseases. Here, the conventional unit for the treatment cost is 1000 (VND). It is noted that, for each attribute $A \in \boldsymbol{U}$={P_ID, P_NAME, AGE, DISEASE D_COST} in the schema $\mathbf{PATIENT}(\boldsymbol{U}, \wp), \wp(A)$ includes all fuzzy probabilistic triples on the value domain of A (Definition 9).

Table 2. Relation PATIENT

P_ID	P_NAME	AGE	DISEASE	D_COST
PT226	*N.V. Ha*	$\langle\{65\}, u, u\rangle$	$\langle\{$*lung cancer*, *tuberculosis*$\}$, $0.8u, 1.2u\rangle$	$\langle\{300, 350\}, u, u\rangle$
PT234	*T.V. Son*	$\langle\{$*young*$\}$,*u*, $u\rangle$	$\langle\{$*hepatitis*, *cirrhosis*$\}$, $0.9u, 1.3u\rangle$	$\langle\{$*about_60*, *about_70*$\}$, $0.8u, 1.2u\rangle$
PT242	*L.T. Lan*	$\langle\{$*middle_aged*$\}$,$\langle\{$*cholecystitis*$\}$, $u, u, u\rangle$		$\langle\{8\}, u, u\rangle$

Also note that, in the real world applications, fuzzy set values of attributes of the fuzzy probabilistic relation PATIENT, such as *about_60*, *about_70*, *young* and *middle_aged* will be defined (maybe by professional experts) compatibly and consistently with the meaning of the information represented by them. In this example, the fuzzy set values only simply illustrate for Definition 10. For example, *about_60* = {58: 0.5, 59: 0.9, 60: 1, 61: 0.9, 62: 0.5} or *about_70* = {68: 0.5, 69: 0.9, 70: 1, 71: 0.9, 72: 0.5} can be considered as a fuzzy set value representing the daily treatment cost of the patient *T.V. Son* who has hepatitis or cirrhosis. Similarly, *young* and *middle_aged* whose membership function graphs as Figure 1, may be considered as fuzzy set values representing the imprecise age of the patients *T.V.Son* and *L.T. Lan* respectively. In addition, for simplicity, each fuzzy probabilistic triple $\langle V, u, u\rangle$, with V={v}, will be represented as a single value v because if the attribute takes such a fuzzy probabilistic triple, then actually it only takes a value v with the probability is 1 (Definition 5). In other words, the attribute certainly takes the value v.

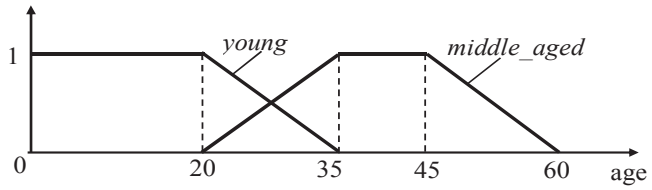

Figure 1. Fuzzy set values of the attribute AGE

Now, the notion of a fuzzy probabilistic relational database is defined as follows.

Definition 11. A *fuzzy probabilistic relational database* over a set of attributes is a set of fuzzy

probabilistic relations corresponding with the set of their fuzzy probabilistic relational schemas.

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

3.3. Fuzzy probabilistic functional dependencies

For defining the fuzzy probabilistic functional dependent concept in FPRDB, first a probability measure is proposed to determine the fuzzy equal degree of two values of the same attribute for two different tuples in a relation as below.

Definition 12. Let t_1 and t_2 be two tuples in a fuzzy probabilistic relation r, A be an attribute of r and $\otimes$ be a probabilistic conjunction strategy. The *probability interval* for the values of the attribute A of two tuples t_1 and t_2 respectively are equal under $\otimes$, 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

$[\alpha(v),\ \beta(v)] = [\alpha_1(v_1),\ \beta_1(v_1)] \otimes [\alpha_2(v_2),\ \beta_2(v_2)]$, $\forall v = (v_1,\ v_2) \in V = V_1 \times V_2$.

Now, the fuzzy probabilistic functional dependency in FPRDB is extended from the probabilistic functional dependency in PRDB [17] as follows.

Definition 13. Let $\boldsymbol{U} = \{A_1, A_2, \ldots, A_k\}$ be a set of k pairwise different attributes $R(\boldsymbol{U},\ \wp)$ be a fuzzy probabilistic relational schema, r be any fuzzy probabilistic relation over R, $\otimes$ be a probabilistic conjunction strategy, $X = \{A_i, \ldots, A_l\}$ and $Y = \{A_j, \ldots, A_m\}$ be two subsets of $\boldsymbol{U}$. A *fuzzy probabilistic functional dependency* of Y on X under $\otimes$ over R, denoted by $X \rightsquigarrow_\otimes Y$, if and only if

$$\forall t_1,\ t_2 \in r,\ \Pr(t_1[X] =_\otimes t_2[X]) \leq \Pr(t_1[Y] =_\otimes t_2[Y]),$$

where

$$\Pr(t_1[X] =_\otimes t_2[X]) = p(t_1.A_i =_\otimes t_2.A_i) \quad \otimes \ldots \otimes p(t_1.A_l =_\otimes t_2.A_l)$$

and

$$\Pr(t_1[Y] =_\otimes t_2[Y]) = p(t_1.A_j =_\otimes t_2.A_j) \quad \otimes \ldots \otimes p(t_1.A_m =_\otimes t_2.A_m).$$

An obvious example of the fuzzy probabilistic functional dependency is every attribute A_i depending on the set $\{A_1, A_2, \ldots,\ A_k\}$ that consists of all attributes of the schema R. It is noted that in the classical database, one can consider the probability for two values of an attribute equal i.e. only 0 or 1, so the functional dependency in the classical relational database is a particular case of the fuzzy probabilistic functional dependency in this definition.

As for the classical relational database model, the keys of a schema in FPRDB are the basis for recognising a tuple in a fuzzy probabilistic relation. In the model and management systems of the classical relational database, key attributes are constrained not to take the value NULL [3, 4]. Similarly, in FPRDB, it is assuming that the value of each key attribute is always certain and definite. The key concept of fuzzy probabilistic relational schema is defined using the fuzzy probabilistic functional dependency as follows.

Definition 14. Let $R(\boldsymbol{U},\ \wp)$ be a fuzzy probabilistic relational schema, r be any relation over R and $\otimes$ be a probabilistic conjunction strategy, a set of attributes $K \subseteq \boldsymbol{U}$ is called a *key of* R under $\otimes$ if the value of each attribute of K is always certain, precise in r and there is a fuzzy probabilistic functional dependency $K \rightsquigarrow_\otimes \boldsymbol{U}$ such that not to exist any subset of K has the properties.

In the relation PATIENT above, if we assume that each patient has a unique identifier corresponding with the value of the attribute P_ID and the identifier differs from every identifier of other patients, then by the definition, P_ID is a key of the schema **PATIENT** under any probabilistic conjunction strategy.

4. SELECTION OPERATION ON A FUZZY PROBABILISTIC RELATION

4.1. Syntax of selection conditions

The selection is a basic algebraic operation and is used to query the imprecise and uncertain information in FPRDB. Before defining the selection operation, we present the formal syntax and semantics of selection conditions by extending those definitions of PRDB with fuzzy sets. We start with the syntax of selection expressions as the following definition.

Definition 15. Let R be a schema in FPRDB and $\boldsymbol{X}$ 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 $x \in \boldsymbol{X}$, A is an attribute in R, θ is a binary relation from $\{=, \neq, \leq, <, \subseteq, \in, \rightarrow\}$ and c is a single value or a fuzzy set.

2) $x.A_1 =_{\otimes} x.A_2$, where $x \in \boldsymbol{X}$, A_1 and A_2 are two different attributes in R and $\otimes$ 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, $\otimes$ 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, $\oplus$ is a probabilistic disjunction strategy.

Example 5. Consider the fuzzy probabilistic relational schema **PATIENT** in Example 4, the selection of "all patients who get hepatitis is and pay the daily treatment cost about 60 (thousand VND)" can be expressed by the selection expression x.DISEASE = *hepatitis*$\otimes x$.D_COST$\rightarrow$*about_60*.

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

Definition 16. Let R be a schema in FPRDB, *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 6. With the schema of the relation PATIENT in Example 4, then the selection of "all patients who are young with a probability of at least 0.4 and have lung cancer with a probability of at least 0.8" can be done using the selection condition (x.AGE $\rightarrow$ young) [0.4, 1.0] $\wedge$ (x.DISEASE = lung cancer)[0.8, 1.0].

4.2. Semantics of selection conditions

For defining the semantics of selection conditions in FPRDB, we first extend probabilistic interpretations of selection expressions in [17] with the binary relations on fuzzy sets in Section 2 as the definition below.

Definition 17. Let R be a fuzzy probabilistic relational schema in FPRDB, 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 $v\theta c$ and $prob_{R,r,t}(x.A_1 =_{\otimes} x.A_2)$ 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 7. Let r denote the relation PATIENT in Example 4 and R denote the schema of PATIENT consider the tuple t_2 (the second tuple) in r, using Definition 3 we have $prob(about\_60 \rightarrow about\_60)$ = 0.86 and $prob(about\_70 \rightarrow about\_60) = 0.0$. From that

$prob_{R,r,t2}(x.\text{D\_COST} \rightarrow about\_60) = [0.8u(about\_60).prob(about\_60 \rightarrow about\_60) +$
$0.8u(about\_70).prob(about\_70 \rightarrow about\_60),$
$\min(1, 1.2u(about\_60).prob(about\_60 \rightarrow about\_60) +$
$1.2u(about\_70).prob(about\_70 \rightarrow about\_60))]$
$= [0.8 \times 0.5 \times 0.86 + 0.8 \times 0.5 \times 0., \min(1, 1.2 \times 0.5 \times 0.86 + 1.2 \times 0.5 \times 0.0)] = [0.344, 0.516]$.

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

Definition 18. Let R be a relational schema in FPRDB, r be a fuzzy probabilistic relation over R and $t \in r$. The *satisfaction* of selection conditions under $prob_{R,r,t}$ is defined as follows:

1) $prob_{R,r,t} \models (E)[L, U]$ if and only if (iff) $prob_{R,r,t}(E) \subseteq [L, U]$.

2) $prob_{R,r,t} \models \neg\phi$ iff $prob_{R,r,t} \models \phi$ does not hold.

3) $prob_{R,r,t} \models \phi \wedge \psi$ iff $prob_{R,r,t} \models \phi$ and $prob_{R,r,t} \models \psi$.

4) $prob_{R,r,t} \models \phi \vee \psi$ iff $prob_{R,r,t} \models \phi$ or $prob_{R,r,t} \models \psi$.

Note that, in the classical database, the concepts of selection expression and selection condition are identical and we can consider probability intervals $[L, U]$ in selection conditions being always equal to [1.0, 1.0]. This also means that the concept of satisfaction of selection conditions in the classical relational database model is a particular case of the concept of satisfaction of selection conditions in FPRDB.

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

Definition 19. Let R be a relational schema in FPRDB, r be a fuzzy probabilistic 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 $r' = \{t \in r | prob_{R,r,t} \models \phi\}$ over R, including all satisfied tuples of the selection condition ϕ.

Example 8. Let *approx_15*= (10: 0; 15: 1; 20: 0) denote the continuous triangle shaped fuzzy set whose vertices are (15, 1), (10, 0) and (20, 0). Consider the relation PATIENT in Example 4. Then, the query "Find all patients who are approximately 15 years old with a probability from 0.2 to 0.5 and have hepatitis with a probability of at least 0.4" can be done by the selection operation r'=σ_ϕ(PATIENT) with ϕ =(x.AGE→*approx_15*)[0.2, 0.5]∧ (x.DISEASE = *hepatitis*)[0.4, 1.0].

As noted in Definition 3, the probabilistic interpretation $prob(A \to B)$ can also be adapted for fuzzy sets on continuous domains, using integration instead of addition. That is

$$prob(A \to B) = \int_0^1 \int_0^1 p(u \in {}^yB | u \in {}^xA) dxdy = \int_0^1 \int_0^1 \frac{\mathrm{p}({}^xA \cap {}^yB)}{\mathrm{p}({}^xA)} dxdy = \int_0^1 \int_0^1 \frac{|{}^xA \cap {}^yB|}{|{}^xA|} dxdy$$

where ${}^xA$ and ${}^yB$ are α -cuts of the fuzzy sets A and B with $\alpha = x$ and $\alpha = y$, respectively.

For the fuzzy sets *approx_15* above and *young* given in Example 4, their α-cuts are respectively ${}^\alpha$*approx_15*= [10+5α, 20-5α] and ${}^\alpha$*young*= [0, 35-15α]. From that $prob(young \to approx\_15)$ is computed as follows:

$$\begin{aligned} prob(young \to approx\_15) &= \int_0^1 \int_0^1 \frac{|{}^x young \cap {}^y approx\_15|}{|{}^x young|} dxdy \\ &= \int_0^1 \int_0^1 \frac{|[0,\, 35-15x] \cap [10+5y,\, 20-5y]|}{|[0,\, 35-15x]|} dxdy \\ &= \int_0^1 \int_0^1 \frac{10-10y}{35-15x} dxdy = \int_0^1 \int_0^1 \frac{2-2y}{7-3x} dxdy = \frac{\ln 7 - \ln 4}{3} \approx 0.2. \end{aligned}$$

The selection is implemented by checking the satisfaction of all tuples in PATIENT for the selection condition ϕ. From the result computed above, we can see only the tuple t_2 satisfies ϕ because $prob_{R,r,t2}(x.\text{AGE}\to approx\_15) = [u(young).prob(young \to approx\_15),\ \min(1,\ u(young).prob(young \to approx\_15))]$ =[1 ×0.2, min(1, 1 ×0.2]= [0.2, 0.2]⊆ [0.2, 0.5] and $prob_{R,r,t2}(x.$ DISEASE = *hepatitis*) = [0.45, 0.65]⊆[0.4, 1.0]. So, the result of the selection is the relation r'= σ_ϕ(PATIENT) as in Table 3.

Table 3. Relation r'=σ_{ϕ} (PATIENT)

P_ID	P_NAME	AGE	DISEASE	D_COST
PT234	*T.V. Son*	$\langle\{young\}$,u, $u\rangle$	$\langle\{hepatitis, cirrhosis\}$, $0.9u$, $1.3u\rangle$	$\langle\{about\_60, about\_70\}$, $0.8u$, $1.2u\rangle$

5. OTHER OPERATIONSON FUZZY PROBABILISTIC RELATIONS

As for the classical relational database and PRDB, other basic operations on fuzzy probabilistic relations are the projection, Cartesian product, join, intersection, union, and difference. We now extend those operations of PRDB for FPRDB taking into account imprecise and uncertain values of relational attributes.

5.1. Projection

A projection of a fuzzy probabilistic relation on a set of attributes is a new fuzzy probabilistic relation in which only the attributes in that set are considered for each tuple of the new relation as the following definition.

Definition 20. Let R= ($\boldsymbol{U}$, $\wp$) be a fuzzy probabilistic relational schema, r be a relation over R and $\boldsymbol{L}$ be a subset of attributes of $\boldsymbol{U}$. The *projection* of r on $\boldsymbol{L}$ denoted by $\Pi_L(r)$, is the fuzzy probabilistic 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' = $\{t$'= $t[\boldsymbol{L}]|t \in r\}$, i.e., r' consists of tuples t' to achieve from the tuples t = ($\langle V_1, \alpha_1, \beta_1\rangle$,..., $\langle V_k, \alpha_k, \beta_k\rangle$) $\in r$ by eliminating every $\langle V_j, \alpha_j, \beta_j\rangle$ that $t.A_j = \langle V_j, \alpha_j, \beta_j\rangle$ and $A_j \notin \boldsymbol{L}$.

5.2. Cartesian product

For the Cartesian product of two fuzzy probabilistic relations, as in the classical relational database and PRDB, we assume the set of attributes of their schemas are disjoint. Also, for the operation being commutative, we assume every k-tuple t = ($\langle V_1, \alpha_1, \beta_1\rangle$,..., $\langle V_k, \alpha_k, \beta_k\rangle$) is an un-ordered list.

Definition 21. The fuzzy probabilistic relational schemas $R_1(\boldsymbol{U}_1, \wp_1)$ and $R_2(\boldsymbol{U}_2, \wp_2)$ are *Cartesian product-compatible* if and only if $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ have not any common attribute.

Note that, any schemas $R_1(\boldsymbol{U}_1, \wp_1)$ and $R_2(\boldsymbol{U}_2, \wp_2)$ can be made Cartesian product-compatible by renaming of attributes in $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$.

Now, the Cartesian product of two fuzzy probabilistic relations in FPRDB is extended from that operation of PRDB as follows.

Definition 22. Let r_1 and r_2 be two fuzzy probabilistic relations over the Cartesian product-compatible schemas R_1= ($\boldsymbol{U}_1, \wp_1$) and R_2= ($\boldsymbol{U}_2, \wp_2$), respectively. The *Cartesian product* of r_1 and r_2, denoted by $r_1 \times r_2$, is the fuzzy probabilistic 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= $\{t$ =($\langle V_1, \alpha_1, \beta_1\rangle, \ldots, \langle V_k, \alpha_k, \beta_k\rangle, \langle V_{k+1}, \alpha_{k+1}, \beta_{k+1}\rangle, \ldots, \langle V_{k+m}, \alpha_{k+m}, \beta_{k+m}\rangle$)$|t_1$ =($\langle V_1, \alpha_1, \beta_1\rangle, \ldots, \langle V_k, \alpha_k, \beta_k\rangle$) and t_2 = ($\langle V_{k+1}, \alpha_{k+1}, \beta_{k+1}\rangle, \ldots, \langle V_{k+m}, \alpha_{k+m}, \beta_{k+m}\rangle$), $t_1 \in r_1$ and $t_2 \in r_2\}$.

5.3. Join

The join of two fuzzy probabilistic relations in FPRDB is extended from the natural join of two relations in the classical relational database and the join in PRDB. The join only works with relations whose schemas are join-compatible as the definition below.

Definition 23. The fuzzy probabilistic relational schemas $R_1(\boldsymbol{U}_1, \wp_1)$ and $R_2(\boldsymbol{U}_2, \wp_2)$ are *join -compatible* if and only if the domains of two attributes of the same name A in $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$, respectively are identical.

From Definition 9, we can see that for two attributes of the same name A in $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ of two join–compatible schemas $R_1(\boldsymbol{U}_1, \wp_1)$ and $R_2(\boldsymbol{U}_2, \wp_2)$ then $\wp_1(A) = \wp_2(A)$. For building the join of fuzzy two probabilistic relations, we first extend the join of two tuples in PRDB for FPRDB using the conjunction of fuzzy probabilistic triples as follows.

Definition 24. Let t_1 and t_2 be two tuples on two sets of attributes $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ respectively, and $\otimes$ be a probabilistic conjunction strategy. The *join* of t_1 and t_2 under $\otimes$, denoted by $t_1 \bowtie_{\otimes} t_2$, is the tuple t on $\boldsymbol{U}_1 \cup \boldsymbol{U}_2$ defined by:

1) $t.A = t_1.A, \forall A \in \boldsymbol{U}_1 - \boldsymbol{U}_2$,

2) $t.A = t_2.A, \forall A \in \boldsymbol{U}_2 - \boldsymbol{U}_1$,

3) $t.A = t_1.A \otimes t_2.A, \forall A \in \boldsymbol{U}_1 \cap \boldsymbol{U}_2$.

It is noted that $t_1.A \otimes t_2.A$ is the conjunction of two fuzzy probabilistic triples $t_1.A$ and $t_2.A$ by Definition 6. It is easy to see that $t_1 \bowtie_{\otimes} t_2 = t_2 \bowtie_{\otimes} t_1$ for every probabilistic conjunction strategy, i.e. the join of two tuples is commutative.

Definition 25. Let r_1 and r_2 be two fuzzy probabilistic relations over the join-compatible schemas R_1= ($\boldsymbol{U}_1$, $\wp_1$) and R_2= ($\boldsymbol{U}_2$, $\wp_2$), respectively and let $\otimes$ 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 fuzzy probabilistic 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 $A \in \boldsymbol{U}_1 \cap \boldsymbol{U}_2$ (because $\wp_1(A) = \wp_2(A)$, Definition 23).

2) r= $\{t = t_1 \bowtie_{\otimes} t_2 |\ t_1 \in r_1, t_2 \in r_2$ and $\forall A \in \boldsymbol{U}_1 \cap \boldsymbol{U}_2$, if $t_1.A \otimes t_2.A = \langle V, \alpha, \beta\rangle$ then $V \neq \emptyset\}$.

Example 9. Given two fuzzy probabilistic relations DOCTOR_1 and DOCTOR_2 as in Tables 4 and 5, where *approx_30*= (25:0;30:1;35:0) denotes the continuous triangle shaped fuzzy set whose vertices are (30, 1), (25, 0) and (35, 0) and *middle_aged* is the fuzzy set 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 (with the height equals 1, by Definition 6) is the relation DOCTOR computed as in Table 6.

Table 4. Relation DOCTOR_1

DOCTOR_ID	D_AGE
DT005	$\langle\{middle\_aged\}, u, u\rangle$
DT093	$\langle\{approx\_30\}, u, u\rangle$
DT102	$\langle\{55,56\}, u, u\rangle$

Table 5. Relation DOCTOR_2

DOCTOR_NAME	D_AGE
L.V Cuong	$\langle\{30, 31\}, 0.8u, 1.2u\rangle$
N.V. Hung	$\langle\{middle\_aged\}, u, u\rangle$
N.T.Dat	$\langle\{54, 55\}, u, u\rangle$

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 $dom(A)$.

5.4. Intersection, union, and difference

Intersection, union and difference of two fuzzy probabilistic relations, respectively, over the same schema is a fuzzy probabilistic relation over that schema, in which the value of attributes in common tuples of those two relations associated by a probabilistic combination strategy. A common tuple of two fuzzy probabilistic relations over the same schema is the tuple whose key attributes' values are identical in both relations. It is due to the impreciseness and uncertainty of attribute values, a common tuple of two fuzzy probabilistic relations is not completely identical as that of two relations in the classical relational database.

First, the intersection of two tuples as the basis for the intersection of two fuzzy probabilistic relations is defined as follows.

Definition 26. Let t_1 and t_2 be two tuples respectively in two fuzzy probabilistic relations over the same schema $R(\boldsymbol{U}, \wp)$ and $\otimes$ be a probabilistic conjunction strategy. The *intersection* of t_1 and t_2 under $\otimes$, denoted by $t_1 \cap_{\otimes} t_2$, is the tuple t on $\boldsymbol{U}$ defined by $t.A = t_1.A \otimes t_2.A$ for every $A \in \boldsymbol{U}$.

Definition 27. Let r_1 and r_2 be two fuzzy probabilistic relations over the same schema $R(\boldsymbol{U}, \wp)$, K be a key of R and $\otimes$ be a probabilistic conjunction strategy. The *intersection* of r_1 and r_2 under $\otimes$, denoted by $r_1 \cap_{\otimes} r_2$, is the fuzzy probabilistic relation r over R defined by $r = \{t = t_1 \cap_{\otimes} t_2 | \quad t_1 \in r_1, t_2 \in r_2$ such that $t_1[K] = t_2[K]$ and $t_1.A \otimes t_2.A \neq \langle\emptyset, \alpha, \beta\rangle, \forall A \in \boldsymbol{U}\}$.

Table 6. Relation DOCTOR= $\text{DOCTOR}_1 \bowtie_{\otimes in} \text{DOCTOR}_2$

DOCTOR_ID	DOCTOR_NAME	D_AGE
DT005	*N.V. Hung*	$\langle\{middle\_aged\}, u, u\rangle$
DT093	*L.V Cuong*	$\langle\{30\}, 0.4u, 0.6u\rangle$
DT102	*N.T.Dat*	$\langle\{55\}, 0.25u, 0.25u\rangle$

It is noted that, the notation $t_1[K] = t_2[K]$ is used in the definition due to the value of each key attribute assumed is certain and definite as in the Definition 14.

Example 10. Consider two relations DIAGNOSE$_1$ and DIAGNOSE$_2$ over the same schema **DIAGNOSE**($\boldsymbol{U}$, $\wp$) as in Tables 7 and 8 where $\boldsymbol{U}$={P_ID, DOCTOR_ ID, P_AGE, DISEASE} and {P_ID, DOCTOR_ ID} is a key of **DIAGNOSE**, *young*, *middle_aged* and *approx_15* are the fuzzy sets given in Examples 4 and 8. 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 $\otimes_{in}$ is the relation DIAGNOSE computed as in Table 9.

Table 7. Relation DIAGNOSE$_1$

P_ID	DOCTOR_ ID	P_AGE	DISEASE
PT226	*DT093*	$\langle\{65\}, u, u\rangle$	$\langle\{$*lung cancer*, *tuberculosis*$\}, 0.8u, 1.2u\rangle$
PT234	*DT102*	$\langle\{$*approx_15*$\}, u, u\rangle$	$\langle\{$*hepatitis*, *cirrhosis*$\}, u, u\rangle$

Table 8. Relation DIAGNOSE$_2$

P_ID	DOCTOR_ID	P_AGE	DISEASE
PT383	*DT102*	$\langle\{69, 70\}, u, u\rangle$	$\langle\{$*lung cancers*$\}, u, u\rangle$
PT234	*DT102*	$\langle\{$*young*$\}, u, u\rangle$	$\langle\{$*hepatitis*, *gall-stone*$\}, 0.8u, 1.2u\rangle$
PT242	*DT025*	$\langle\{$*middle_aged*$\}, u, u\rangle$	$\langle\{$*cholecystitis*$\}, u, u\rangle$

Table 9. Relation DIAGNOSE = DIAGNOSE$_1 \cap_{\otimes in}$DIAGNOSE$_2$

P_ID	DOCTOR_ID	P_AGE	DISEASE
PT234	*DT102*	$\langle\{$*approx_15*$\}, u, u\rangle$	$\langle\{$*hepatitis*$\}, 0.2u, 0.3u\rangle$

It is noted that, in the example, the *standard intersection* is used to compute the intersection of fuzzy sets whereby, for example *approx_15* $\cap$ *young*= *approx_15*. In addition, by Definition 6, the height of each fuzzy set computed from the intersection of two fuzzy sets is required to equal 1.

The union of two fuzzy probabilistic relations over the same schema in FPRDB are based on the union of tuples as below.

Definition 28. Let t_1 and t_2 be two tuples respectively in two fuzzy probabilistic relations over the same schema $R(\boldsymbol{U}, \wp)$ and $\oplus$ be a probabilistic disjunction strategy. The *union* of t_1 and t_2 under $\oplus$, denoted by $t_1 \cup_\oplus t_2$, is the tuple t on $\boldsymbol{U}$ defined by $t.A = t_1.A \oplus t_2.A$ for every $A \in \boldsymbol{U}$.

Definition 29. Let r_1 and r_2 be two fuzzy probabilistic relations over the same schema $R(\boldsymbol{U}, \wp)$, K be a key of R and $\oplus$ be a probabilistic disjunction strategy. The *union* of r_1 and r_2 under $\oplus$, denoted by $r_1 \cup_\oplus r_2$, is the fuzzy probabilistic relation r over R defined by $r = \{t_1 \in r_1|$ there is not any tuple $t_2 \in r_2$ such that $t_1[K] = t_2[K]\} \cup \{t_2 \in r_2|$ there is not any tuple $t_1 \in r_1$ such that $t_2[K] = t_1[K]\} \cup \{t = t_1 \cup_\oplus t_2 | t_1 \in r_1, t_2 \in r_2$ such that $t_1[K] = t_2[K]\}$.

As for the intersection and union operations, for defining the difference operation of two fuzzy probabilistic relations, we first define the difference operation of two tuples as follows.

Definition 30. Let t_1 and t_2 be two tuples respectively in two fuzzy probabilistic relations over the same schema $R(\boldsymbol{U}, \wp)$ and $\ominus$ be a probabilistic difference strategy. The *difference* of t_1 and t_2 under $\ominus$, denoted by $t_1 -_{\ominus} t_2$, is the tuple t on $\boldsymbol{U}$ defined by $t.A = t_1.A \ominus t_2.A$ for every $A \in \boldsymbol{U}$.

Definition 31. Let r_1 and r_2 be two fuzzy probabilistic relations over the same schema $R(\boldsymbol{U}, \wp)$, K be a key of R and $\ominus$ be a probabilistic difference strategy. The *difference* of r_1 and r_2 under $\ominus$, denoted by $r_1 -_{\ominus} r_2$, is the fuzzy probabilistic relation r over R defined by $r= \{t_1 \in r_1 |$ there is not any tuple $t_2 \in r_2$ such that $t_1[K] = t_2[K]\} \cup \{t = t_1 -_{\ominus} t_2 |\ \ t_1 \in r_1, t_2 \in r_2$ such that $t_1[K] = t_2[K]$ and $t_1.A \ominus t_2.A \neq \langle \emptyset, \alpha, \beta \rangle, \forall A \in \boldsymbol{U}\}$.

6. PROPERTY OF ALGEBRAIC OPERATIONS

In this section, we propose some properties of the fuzzy probabilistic relational algebraic operations in FPRDB as an extension from those in the classical relational database and PRDB. Clearly, these properties say that FPRDB model is coherent and consistent.

Theorem 1. *Let r be a fuzzy probabilistic relation over the schema R in FPRDB, ϕ_1 and ϕ_2 be two selection conditions. Then*

$$\sigma_{\phi 1}(\sigma_{\phi 2}(r)) = \sigma_{\phi 2}(\sigma_{\phi 1}(r)) = \sigma_{\phi 1 \wedge \phi 2}(r) \tag{1}$$

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

The first property shows that the selections may be reordered.

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

$$\begin{aligned}
\sigma_{\phi 1}(\sigma_{\phi 2}(r)) &= \{t \in r_2 | prob_{R,r2,t} \models \phi_1\} \\
&= \{t \in r | (prob_{R,r,t} \models \phi_2) \quad \wedge (prob_{R,r2,t} \models \phi_1)\} \\
&= \{t \in r | (prob_{R,r,t} \models \phi_2) \quad \wedge (prob_{R,r,t} \models \phi_1)\} (\text{because of } r_2 \subseteq r) \\
&= \{t \in r | prob_{R,r,t} \models \phi_1 \wedge \phi_2\} \text{ (Definition 18)} \\
&= \sigma_{\phi 1 \wedge \phi 2}(r).
\end{aligned}$$

So, $\sigma_{\phi 1}(\sigma_{\phi 2}(r)) = \sigma_{\phi 1 \wedge \phi 2}(r)$ is proven. 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, Theorem 1 is proven. ∎

Theorem 2. *Let r be a fuzzy probabilistic relation over the schema R in FPRDB, $\boldsymbol{A}$ and $\boldsymbol{B}$ be two subsets of attributes of R and $\boldsymbol{A} \subseteq \boldsymbol{B}$. Then*

$$\Pi_A(\Pi_B(r)) = \Pi_A(r). \tag{2}$$

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 20). From that, we are easy to see $\Pi_A(\Pi_B(r)) = \Pi_{A\cap B}(r) = \Pi_A(r)$. Thus, equation(2) is proven. ∎

Theorem 3. *Let R_1, R_2 and R_3 be pairwise join-compatible schemas in FPRDB, r_1, r_2 and r_3 be fuzzy probabilistic relations over R_1, R_2 and R_3 respectively. Let $\otimes$ be a probabilistic conjunction strategy. Then*

$$r_1 \bowtie_{\otimes} r_2 = r_2 \bowtie_{\otimes} r_1, \tag{3}$$

$$(r_1 \bowtie_{\otimes} r_2) \bowtie_{\otimes} r_3 = r_1 \bowtie_{\otimes} (r_2 \bowtie_{\otimes} r_3). \tag{4}$$

Equation (4) and (5) say that the join operation of fuzzy probabilistic 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 6, the conjunction of fuzzy probabilistic triples is commutative (due to the commutativity of probabilistic conjunction strategies and the intersection of fuzzy sets). Consequently, the join of tuples is commutative (by Definition 24). So, by Definition 25, we have $r_1 \bowtie_{\otimes} r_2 = r_2 \bowtie_{\otimes} r_1$.

Since R_1, R_2 and R_3 are pairwise join-compatible, so 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 6, it follows that the conjunction of fuzzy probabilistic triples is associative. From associativity of the classical relational join and by Definition 24, it is easy to see that the join of tuples which are based on the conjunction of fuzzy probabilistic triples is associative. Thus, by Definition 25, 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 22 and Definition 25), we have the straight corollary of Theorem 3 below.

Corollary 1. *Let R_1, R_2 and R_3 be pairwise Cartesian product-compatible schemas in FPRDB, r_1, r_2 and r_3 be fuzzy probabilistic relations over R_1, R_2 and R_3 respectively. Then*

$$r_1 \times r_2 = r_2 \times r_1, \tag{5}$$

$$(r_1 \times r_2) \times r_3 = r_1 \times (r_2 \times r_3). \tag{6}$$

Theorem 4. *Let r_1, r_2 and r_3 be fuzzy probabilistic relations over the same schema R in FPRDB. Let $\otimes/\oplus$ be a probabilistic conjunction/disjunction strategy. Then*

$$r_1 \cap_{\otimes} r_2 = r_2 \cap_{\otimes} r_1, \tag{7}$$

$$(r_1 \cap_{\otimes} r_2) \cap_{\otimes} r_3 = r_1 \cap_{\otimes} (r_2 \cap_{\otimes} r_3), \tag{8}$$

$$r_1 \cup_{\oplus} r_2 = r_2 \cup_{\oplus} r_1, \tag{9}$$

$$(r_1 \cup_{\oplus} r_2) \cup_{\oplus} r_3 = r_1 \cup_{\oplus} (r_2 \cup_{\oplus} r_3). \tag{10}$$

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

Proof. Equations in the theorem are proven respectively as follows:

Equations (7) and (8): From commutativity and associativity of the intersection of fuzzy sets, it follows the conjunction of fuzzy probabilistic triples has commutativity and associativity. Thus, the intersection of tuples, by Definition 26, has commutativity and associativity. So, the intersection of tuples that have the same key value in r_1, r_2 and r_3 respectively is commutative and associative. From that, it follows Equations (7) and (8).

Equations (9) and (10): From commutativity of the union, intersection of fuzzy sets, the disjunction of fuzzy probabilistic triples (Definition 7) and the union of tuples (Definition 28), by Definition 29 we have Equation (9).

For Equation (10), let K be the key used to determine common tuples of r_1, r_2 and r_3. Without loss of generality, we may assume that each tuple t belongs to one of the three relations r_1, r_2 and r_3 then there exists two tuples belonging to the two remaining relations, respectively such that the value of the key K of the three tuples respectively are always the same. This can be done by adding t to the relations in which it is missing and resetting $\alpha(v) = \beta(v) = 0$ for every v in the set V of values of each attribute $A \notin K$ of the tuple t. Under this technical assumption, the result of each expression in Equation (10) is not changed and only the union case of two tuples in Definition 29 is relevant. Now, from associativity of the disjunction of fuzzy probabilistic triples (Definition 7) and the union of tuples, Equation(10) obviously holds. ■

7. CONCLUSION

In this paper, the authors propose a fuzzy probabilistic relational database model, called FPRDB, as an extension of the PRDB model with fuzzy sets. FPRDB is also a complete development for the model in [18] with a full set of basic fuzzy probabilistic relational algebraic operations. FPRDB is capable of representing and manipulating both imprecise and uncertain information in the real world applications. FPRDB has been built based on the association of the probability theory and fuzzy theory. The relational schemas, relations, functional dependencies and relational algebraic operations of FPRDB have been defined coherently and consistently. A set of basic properties of the algebraic operations in FPRDB has also been proposed as theorems and proven completely.

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

REFERENCES

[1] T. Eiter, J. J. Lu, T. Lukasiewicz, and V. Subrahmanian, "Probabilistic object bases," *ACM Transactions on Database Systems (TODS)*, vol. 26, no. 3, pp. 264–312, 2001.

[2] 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, no. 02, pp. 275–305, 2011.

[3] E. F. Codd, "A relational model of data for large shared data banks," *Communications of the ACM*, vol. 13, no. 6, pp. 377–387, 1970.

[4] D. C. J., *An introduction to database systems.* Prentice hall New Jersey, 8th ed., 2008.

[5] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic.* Prentice hall New Jersey, 1995, vol. 4.

[6] Z. L.A., "Fuzzy sets," *Journal of Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[7] T. Ge, A. Dekhtyar, and J. Goldsmith, "Uncertain data: Representations, query processing, and applications," in *Advances in Probabilistic Databases for Uncertain Information Management*. Springer, 2013, pp. 67–108.

[8] Z. Ma and L. Yan, *Advances in probabilistic databases for uncertain information management*. Springer, 2013, vol. 304.

[9] N. Hoa and T. D. Hieu, "A probabilistic relational data model for uncertain information," in *2013 IEEE Third International Conference on Information Science and Technology (ICIST)*. IEEE, 2013, pp. 607–613.

[10] R. Ross, V. Subrahmanian, and J. Grant, "Aggregate operators in probabilistic databases," *Journal of the ACM (JACM)*, vol. 52, no. 1, pp. 54–101, 2005.

[11] W. Zhao, A. Dekhtyar, and J. Goldsmith, "Databases for interval probabilities," *International journal of intelligent systems*, vol. 19, no. 9, pp. 789–815, 2004.

[12] X. Meng, Z. Ma, and X. Zhu, "A knowledge-based fuzzy query and results ranking approach for relational databases," *Journal of Computational Information Systems*, vol. 6, no. 6, 2010.

[13] J. Mishra and S. Ghosh, "A new functional dependency in a vague relational database model," *International Journal of Computer Applications*, vol. 39, no. 8, pp. 29–36, 2012.

[14] F. Petry, "Fuzzy databases: Principles and applications (with chapter contribution by patrick bosc). international series in intelligent technologies. ed. hj zimmermann," *J. Zimmermann. Kluwer Academic Publishers (KAP)*, 1996.

[15] A. A. Sabour, A. M. Gadallah, and H. A. Hefny, "Flexible querying of relational databases: Fuzzy set based approach," in *International Conference on Advanced Machine Learning Technologies and Applications*. Springer, 2014, pp. 446–455.

[16] L. Yan and Z. Ma, "A fuzzy probabilistic relational database model and algebra," *International Journal of Fuzzy Systems*, vol. 15, no. 1, p. 244, 2013.

[17] H. Nguyen, "A probabilistic relational database model and algebra," *Journal of Computer Science and Cybernetics*, vol. 31, no. 4, p. 305, 2015.

[18] ——, "Towards a fuzzy probabilistic relational data base model," in *Knowledge and Systems Engineering (KSE), 2015 Seventh International Conference on*. IEEE, 2015, pp. 298–301.

[19] J. Baldwin, J. Lawry, and T. Martin, "A mass assignment theory of the probability of fuzzy events," *Fuzzy Sets and Systems*, vol. 83, no. 3, pp. 353–367, 1996.

[20] ——, "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*, 1996, pp. 521–525.

Received on March 15 - 2016
Revised on August 18 - 2016