A Type-2 Fuzzy Relational Database Model

Nguyen Hoa
Saigon University, Ho Chi Minh City, Vietnam
E-mail: nguyenhoa@sgu.edu.vn
Communication: received 11 January 2017, revised 17 July 2017, accepted 31 July 2017

Abstract: This paper introduces a type-2 fuzzy relational database model (T-2FRDB) as an extension of type-1 fuzzy relational database models with a full set of basic fuzzy relational algebraic operations that can represent and query uncertain and imprecise information in real world applications. In this model, the membership degree of tuples in a fuzzy relation is represented by fuzzy numbers on $[0, 1]$, and fuzzy relational algebraic operations are defined by using the extension principle for computing minimum and maximum values of such fuzzy numbers. Some properties of the type-2 fuzzy relational algebraic operations in T-2FRDB are also formulated and proven as extensions of their counterpart in the type-1 fuzzy relational database models.

Keywords: *Fuzzy set, fuzzy relation, type-2 fuzzy relational database, type-2 fuzzy relational algebraic operation.*

I. Introduction

As we know, the classical relational database model [1] is very useful for modeling, designing and implementing large-scale systems. However, it is restricted to representing and handling uncertain and imprecise information about objects in practice [1, 2]. For example, applications of the classical relational database model can not deal with a query like "find all patients who are *young* and have lung cancer and a *high* treatment cost", where *young* and *high* are the vague notion and imprecise value [3, 4].

So far, there have been many relational database models studied and built based on the fuzzy theory for modeling objects about which information may be vague and imprecise to overcome the limitations of the classical relational database model such as [5–11]. Such models are called *fuzzy relational database models*.

There are two main approaches to represent fuzzy relations in fuzzy relational database models. The first approach represents each fuzzy relation as a set of tuples whose each attribute may take a fuzzy set (or a possibility distribution inferred from a fuzzy set) [5–7, 12–14], whereby the membership degree of tuples for the relation is hidden in that of their attribute values. The second one represents each fuzzy relation as a fuzzy set of tuples whose each attribute only takes a single and precise value [3, 8–11, 15–18], whereby the membership degree of tuples for the relation also is that for the fuzzy set.

Fuzzy relational database models that are built based on one of above approaches are extensions of the classical relational database model with fuzzy sets. They have different capabilities for expressing and dealing with uncertain and imprecise information.

There are two types of the models based on the second approach, including the *type-1 fuzzy relational database model* (T-1FRDB), or the (ordinary) fuzzy relational database model, whereby the membership degree of tuples is assigned to a number in $[0, 1]$, and the *type-2 fuzzy relational database model* (T-2FRDB), whereby the membership degree of tuples is expressed as a fuzzy number on $[0, 1]$. Many type-1 fuzzy relational database models have been proposed such as [3, 8–10, 17, 18]. However, since the membership degree of tuples is expressed as a number in $[0, 1]$, these models were restricted in representing the associated imprecise degree of attribute values. In real world relational databases, since attribute values of tuples may be imprecise, there are many situations in which we do not know exactly the membership degree of tuples as a number in $[0, 1]$ but only can estimate it as an approximate number (or a fuzzy number) on $[0, 1]$. Some type-2 fuzzy relational database models have been introduced to overcome the shortcoming of type-1 fuzzy relational database models [11, 15, 16, 19]. However, in [19], only notions of the relational schema and instance were defined but relational algebraic operations were not introduced. In [16], data representative notions were not formally defined and some fuzzy relational algebraic operations were missing. Also, in [11, 15], the set of basic fuzzy relational algebraic operations was not complete. Thus, the abilities to express and deal with imprecise information of those models were limited.

In this paper, we propose a type-2 fuzzy relational database model (T-2FRDB) as an extension of the type-1 fuzzy relational database models to overcome the shortcomings of the models in [11, 15, 16]. In our T-2FRDB, the notions of the data representative model are completely defined formally, the full set of basic type-2 fuzzy relational algebraic operations is built based on the extension principle for computing the minimum and maximum values of fuzzy numbers. Some properties of these algebraic operations are also formulated as theorems and proven coherently. T-2FRDB allows representation of soft queries associated with fuzzy sets to handle imprecise information in practice.

The mathematics that we used to develop T-2FRDB is presented in Section II. Schemas and relations of T-2FRDB are introduced in Section III. Type-2 fuzzy relational algebraic operations and their properties in T-2FRDB are presented in Sections IV and V, respectively. Finally, conclusions and future research directions are given in Section VI.

II. Fuzzy Sets and Fuzzy Relations

In this section, we present some notions about fuzzy sets and fuzzy relations as a mathematical basis for developing the T-2FRDB model. Fuzzy sets are used to represent and execute soft queries while relations in T-2FRDB are defined by fuzzy relations.

1. Fuzzy Sets

For a classical set, an element is to be or not to be in the set or, in other words, the membership degree of an element in the set is binary. For a fuzzy set, the membership degree of an element in the set is expressed by a real number in the interval $[0, 1]$. The fuzzy set is extended from the classical set as in [4] and is defined as follows.

Definition 1: A **fuzzy set** A on a domain $\mathcal{X}$ is defined by a membership function μ_A from $\mathcal{X}$ to the closed interval $[0, 1]$. For each $x \in \mathcal{X}$, $\mu_A(x)$ is the membership degree of x for A.

We note that a classical set A on $\mathcal{X}$ is also a fuzzy set [3] with the membership function $\mu_A(x) = 1$ for all $x \in A$ and $\mu_A(x) = 0$ for all $x \notin A$. Even an element e in $\mathcal{X}$ is also considered as a special fuzzy set on $\mathcal{X}$ with the membership function $\mu_e(e) = 1$ and $\mu_e(x) = 0$, for all $x \in \mathcal{X}$ and $x \neq e$. The fuzzy set A on $\mathcal{X}$ as in the above definition is called the **ordinary fuzzy set** and can be denoted by $A = \{x\colon \mu_A(x) \mid x \in \mathcal{X}\}$. In addition, the notation $A(x)$ can be used to replace $\mu_A(x)$.

The **support** of a fuzzy set A on $\mathcal{X}$ is a classical set containing all elements of $\mathcal{X}$ that have nonzero membership degrees in A. The **height** $h(A)$ of a fuzzy set A on $\mathcal{X}$ is the largest membership degree obtained from all elements in the set. It means that $h(A) = \sup_{x\in\mathcal{X}} A(x)$. A fuzzy set A is said to be **normal** when $h(A) = 1$ and **subnormal** when $h(A) < 1$. A fuzzy set A on the set of real number $\mathbb{R}$ is said to be **convex** if for any elements x, y, z in the support of A, the relation $x < y < z$ implies that $\mu_A(y) \geq \min(\mu_A(x), \mu_A(z))$.

Operations on fuzzy sets are generally defined based on functions from the Cartesian product of closed intervals $[0, 1]$ to the closed interval $[0, 1]$. However, the section only presents standard operations in [4] which are applied in computing the relations of T-2FRDB.

Definition 2: Let A and B be two fuzzy sets on $\mathcal{X}$ and have the membership functions μ_A and μ_B, respectively. The **complement** of A, **union**, **intersection** and **difference** of A and B are defined by their membership functions, for all $x \in \mathcal{X}$, as follows:

1) $\mu_{A^c}(x) = 1 - \mu_A(x)$;
2) $\mu_{A\cup B}(x) = \max(\mu_A(x), \mu_B(x))$;
3) $\mu_{A\cap B}(x) = \min(\mu_A(x), \mu_B(x))$;
4) $\mu_{A-B}(x) = \min(\mu_A(x), 1 - \mu_B(x))$.

Fuzzy numbers are special fuzzy sets that are used to represent the fuzzy relations in T-2FRDB, and were defined in [20], as follows.

Definition 3: A **fuzzy number** A is a fuzzy set on the set of real number $\mathbb{R}$ such that

1) A is a normal and convex fuzzy set;
2) The support of A is bounded.

Example 1: The fuzzy set *about*_0.5, given by a membership function and its graph as shown in Figure 1, is a fuzzy number.

$$about\_0.5(x) = \begin{cases} 2x, & \forall x \in [0, 0.5], \\ 2(1-x), & \forall x \in (0.5, 1], \\ 0, & \forall x \notin [0, 1]. \end{cases}$$

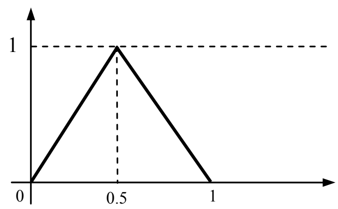

Figure 1. Fuzzy number *about*_0.5.

For computing and combining the membership degrees of tuples in type-2 fuzzy relational algebraic operations, we use two operations MIN and MAX, defined by using the extension principle in [3], as follows.

Definition 4: Let A and B be two fuzzy numbers. The **minimum value** and **maximum value** of A and B are fuzzy numbers that are defined for all $z \in \mathbb{R}$ by

1) $\mathrm{MIN}(A, B)(z) = \sup_{z=\min(x,y)} \min[A(x), B(y)]$;
2) $\mathrm{MAX}(A, B)(z) = \sup_{z=\max(x,y)} \min[A(x), B(y)]$.

Example 2: Let $A = \{1:1, 0.9:0.8, 0.8:0.3\}$ and $B = \{0.6:0.3, 0.5:1, 0.4:0.4\}$ be two fuzzy numbers, then $\mathrm{MIN}(A, B) = \{0.6\!:0.3, 0.5\!:\!1, 0.4\!:\!04\}$.

The fuzzy set of type 2 is extended from the ordinary fuzzy set in [3] as below.

Definition 5: Let $\Im([0,1])$ be the set of all ordinary fuzzy sets on $[0,1]$. A type-2 fuzzy set A on a domain $\mathcal{X}$ is defined by a membership function μ_A from $\mathcal{X}$ to $\Im([0,1])$. For each $x \in \mathcal{X}$, $\mu_A(x)$ is the membership degree of x for A.

We note that since each number in $[0,1]$ is considered as a special ordinary fuzzy set, each ordinary fuzzy set is also considered as a special type-2 fuzzy set.

2. Fuzzy Relations

The ordinary fuzzy relation and type-2 fuzzy relation are the foundation for fuzzy relational database models. The fuzzy relation and type-2 fuzzy relation are defined in [21] by extending the notion of the classical relation based on fuzzy sets as follows.

Definition 6: Let $A_1, A_2, \ldots, A_k$ be non-empty sets. A **k-ary ordinary fuzzy relation** R on these sets is an ordinary fuzzy set on the Cartesian product $A_1 \times A_2 \times \cdots \times A_k$.

Definition 7: Let $A_1, A_2, \ldots, A_k$ be non-empty sets. A **k-ary type-2 fuzzy relation** R on these sets is a type-2 fuzzy set on the Cartesian product $A_1 \times A_2 \times \cdots \times A_k$.

We note that the ordinary fuzzy relation is also called a **type-1 fuzzy relation**. The membership functions of the ordinary fuzzy relation and type-2 fuzzy relation are $\mu_R : A_1 \times A_2 \times \cdots \times A_k \longrightarrow [0,1]$ and $\mu_R : A_1 \times A_2 \times \cdots \times A_k \longrightarrow \Im([0,1])$, respectively.

III. T-2FRDB Schemas and Relations

1. Type-2 fuzzy Relational Schemas

In the following, we will define a special schema in T-2FRDB, which consists of a set of attributes associated with a membership function of a type-2 fuzzy set that is used as a basis for determining type-2 fuzzy relations.

Definition 8: A **type-2 fuzzy relational schema** is a pair $R = (\mathbf{U}, \mu)$, where

1) $\mathbf{U} = \{A_1, A_2, \ldots, A_k\}$ is a set of pairwise different attributes;

2) μ is a function that maps each $(v_1, v_2, \ldots, v_k) \in D_1 \times D_2 \times \cdots \times D_k$ to a fuzzy number on $[0,1]$, where D_i are the domains of the attributes A_i $(i = 1, \ldots, k)$.

As in the classical relational database model, the notations $R(\mathbf{U}, \mu)$ and R can be used to replace $R = (\mathbf{U}, \mu)$. In addition, each $t = (v_1, v_2, \ldots, v_k)$ is called a tuple on the set of attributes $\{A_1, A_2, \ldots, A_k\}$.

TABLE I
Relation PATIENT

P_NAME	AGE	DISEASE	D_COST	μ
L. V. A	53	Lung cancer	350	0.9
L. T. B	65	Cirrhosis	40	*about*_0.5
N. T. C	29	Bronchitis	70	1.0
T. T. D	21	Hepatitis	30	high

Example 3: A type-2 fuzzy relational schema PATIENT in T-2FRDB describing patients can be given as

$$\mathrm{PATIENT(P\_NAME, AGE, DISEASE, D\_COST}, \mu),$$

where μ : string $\times$ integer $\times$ string $\times$ real $\longrightarrow \wp([0,1])$; $\wp([0,1])$ is the set of all fuzzy numbers on $[0,1]$, string, integer and real are domains of the attributes P_NAME, DISEASE, AGE and D_COST, respectively.

2. Type-2 Fuzzy Relations

The following definition extends the notion of ordinary fuzzy relation in T-1FRDB to T-2FRDB.

Definition 9: Let $\mathbf{U} = \{A_1, A_2, \ldots, A_k\}$ be a set of k pairwise different attributes. A **type-2 fuzzy relation** r over the type-2 fuzzy relational schema $R(\mathbf{U}, \mu)$, is a finite set of tuples $\{t_1, t_2, \ldots, t_n\}$ on the set of $\{A_1, A_2, \ldots, A_k\}$ in which each tuple t_i is associated with the fuzzy number $\mu(t_i)$ representing the membership degree of t_i in r, for every $i = 1, 2, \ldots, k$. The notation $t.A$ or $t[A]$ is used to denote the value of attribute A of tuple t in r. The membership degree of t_i in r is denoted by $\mu_r(t_i)$.

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.

Note that, as in the classical database model, if we only care about a unique relation over a schema, we can unify its symbol name with its schema's name.

Example 4: A type-2 fuzzy relation over the schema PATIENT in Example 3 can be given as shown in Table I. In the relation, the attributes P_NAME, AGE, DISEASE, and D_COST provide information about name, age, disease and daily treatment cost of each patient, respectively. In reality, the disease of each patient is not always exactly determined by physicians. Similarly, the daily treatment cost for patients is also not known even as the patients know about their diseases. Here, the conventional unit for the treatment cost is 1000 VND.

We note that $\mu(t)$ represents the membership degree of each tuple t in the relation (by Definition 9). It means that the precise degree of information about the attribute values is expressed by t. For example, let consider the first tuple t_1 in the relation PATIENT and assume that information about the patient's name (L. V. A) expressed by t_1 is correct. The $\mu(t_1) = 0.9$ of t_1 represents the aggregated precise degree of information about the age (53), diagnosed disease (Lung cancer) and daily treatment cost (350,000 VND) of the patient. With t_4, we do not know precisely both information about the attribute values that it represents and its membership degree in the relation PATIENT. We are able to just estimate that $\mu(t_4)$ is *high* where $high = \{0.5{:}0, 0.6{:}0.5, 0.7{:}0.8, 0.8{:}0.9, 0.9{:}1.0, 1{:}1.0\}$ is a fuzzy number on $[0, 1]$.

In real world applications, fuzzy sets that represent the membership degrees of tuples in a fuzzy relation, like *high* and *about*_0.5 as mentioned above, will be defined adequately and consistently based on the meaning and precise degree of information about that these tuples express. The fuzzy sets *high* and *about*_0.5 given in this example are simply meant to give illustration for Definition 9.

Definition 10: A **type-2 fuzzy relational database** over a set of attributes is a set of type-2 fuzzy relations corresponding with the set of their type-2 fuzzy relational schemas.

We note that when $\mu(t) \in [0, 1]$ for every tuple t in a type-2 fuzzy relation, this relation becomes a type-1 fuzzy relation. In other words, a type-1 fuzzy relational database is a particular case of an T-2 FRDB by Definition 10.

IV. Selection Operation on T-2FRDB

1. Syntax of Selection Conditions

The syntax of selection conditions in T-2FRDB is extended from those in [17] with type-2 fuzzy relations as the following definition.

Definition 11: Let R be a schema in T-2FRDB, $\mathcal{X}$ be a set of relational tuple variables and θ be a binary relation from $\{=, \neq, \leq, <, >, \geq\}$. Then **selection conditions** are inductively defined and have one of the following forms:

1) $x.A\ \theta\ v$, where $x \in \mathcal{X}$, A is an attribute in R and v a precise value;

2) $x.A \rightarrow v$, where $x \in \mathcal{X}$, A is an attribute in R, $\rightarrow$ a binary fuzzy relation and v a fuzzy set value;

3) $x.A_1\ \theta\ x.A_2$, where $x \in \mathcal{X}$, A_1 and A_2 are two different attributes in R;

4) $\neg E$, if E is a selection condition;

5) $E_1 \wedge E_2$, if E_1 and E_2 are selection conditions on the same relational tuple variable;

6) $E_1 \vee E_2$, if E_1 and E_2 are selection conditions on the same relational tuple variable.

Example 5: Consider the schema PATIENT in Example 4, the selection of "all patients who are *young* and diagnosed *hepatitis*" can be expressed by the selection condition $x.\text{AGE} \rightarrow young \wedge x.\text{DISEASE} = hepatitis$.

2. Semantics of Selection Conditions

The semantics of selection conditions is the satisfied degree measure for tuples in a fuzzy relation and is defined as follows.

Definition 12: Let $R(\mathbf{U}, \mu)$ be a fuzzy relational schema in T-2FRDB, r a relation over R, x be a tuple variable and t a tuple in r. The **interpretation** of selection conditions with respect to R, r and t, denoted by $\text{Int}_{R,r,t}$, is a partial mapping from the set of all selection conditions to the set of all fuzzy numbers on $[0, 1]$ that is inductively defined as follows:

1) $\text{Int}_{R,r,t}(x.A\ \theta\ v) = \mu_r(t)$ if $t.A\ \theta\ v$,
and $\text{Int}_{R,r,t}(x.A\ \theta\ v) = 0$, otherwise;

2) $\text{Int}_{R,r,t}(x.A \rightarrow v) = \text{MIN}(\mu_r(t), \mu_\varphi(t))$,
with $\varphi = x.A \rightarrow v$;

3) $\text{Int}_{R,r,t}(x.A_1\ \theta\ x.A_2) = \mu_r(t)$ if $t.A_1\ \theta\ t.A_2$,
and $\text{Int}_{R,r,t}(x.A_1\ \theta\ x.A_2) = 0$, otherwise;

4) $\text{Int}_{R,r,t}(\neg E) = 1 - \text{Int}_{R,r,t}(E)$;

5) $\text{Int}_{R,r,t}(E_1 \wedge E_2) = \text{MIN}(\text{Int}_{R,r,t}(E_1), \text{Int}_{R,r,t}(E_2))$;

6) $\text{Int}_{R,r,t}(E_1 \vee E_2) = \text{MAX}(\text{Int}_{R,r,t}(E_1), \text{Int}_{R,r,t}(E_2))$.

We note that v is a fuzzy set in $x.A \rightarrow v$, so $\varphi = x.A \rightarrow v$ is a binary fuzzy relation. Consequently, φ is also a fuzzy set. In particular, φ is the fuzzy set whose elements are tuples in r. For each $t \in r$, $\mu_\varphi(t) = v(t.A)$.

Intuitively, $\text{Int}_{R,r,t}(x.A\ \theta\ v)$ and $\text{Int}_{R,r,t}(x.A \rightarrow v)$ are respectively the satisfied degrees of the conditions $t.A\ \theta\ v$ and $t.A \rightarrow v$ for the tuple t in r while $\text{Int}_{R,r,t}(x.A_1\ \theta\ x.A_2)$ is the satisfied degree of the condition $t.A_1\ \theta\ t.A_2$ for the tuple t in r.

In the classical relational database model, for each tuple t and a relation r, $\mu_r(t) \in \{0, 1\}$, so the interpretation of selection conditions with respect to r and t always takes one of two values 0 or 1. It also means that the concept of interpretation of selection conditions in the classical relational database model is a particular case of that in type-1 fuzzy relational database models and T-2FRDB.

Example 6: Let the fuzzy set *young* represent the young age of a patient whose membership function is defined by

$$\mu_{young}(x) = \begin{cases} 1, & \forall x \in [0, 20], \\ (35 - x)/15, & \forall x \in (20, 35), \\ 0, & \forall x \geq 35. \end{cases}$$

The interpretation of selection conditions

E_1 = "$x.\text{AGE} \rightarrow young$",

E_2 = "$x.\text{DISEASE} = hepatitis$",

E = "$x.\text{AGE} \rightarrow young \wedge x.\text{DISEASE} = hepatitis$", with respect to r = PATIENT and t_4 (the fourth tuple) in Example 4, are

$$\begin{aligned}
&\text{Int}_{R,r,t_4}(E_1)\\
&\quad = \text{MIN}(\mu_r(t_4), young(21)) = \text{MIN}(high, 0.93)\\
&\quad = \{0.5:0, 0.6:0.5, 0.7:0.8, 0.8:0.9, 0.9:1.0, 0.93:1.0\},\\
&\text{Int}_{R,r,t_4}(E_2) = \mu_r(t_4) = high,\\
&\text{Int}_{R,r,t_4}(E)\\
&\quad = \text{MIN}(\text{Int}_{R,r,t_4}(E_1), \text{Int}_{R,r,t_4}(E_2))\\
&\quad = \text{MIN}(\{0.5:0, 0.6:0.5, 0.7:0.8, 0.8:0.9,\\
&\qquad\qquad 0.9:1.0, 0.93:1.0\}, high)\\
&\quad = \{0.5:0, 0.6:0.5, 0.7:0.8, 0.8:0.9, 0.9:1.0, 0.93:1.0\}.
\end{aligned}$$

Let $approx\_0.9 = \{0.5:0, 0.6:0.5, 0.7:0.8, 0.8:0.9, 0.9:1.0, 0.93:1.0\}$. We then have $\text{Int}_{R,r,t_4}(E) = approx\_0.9$.

Now, the selection operation in T-2FRDB is extended from the selection operation in [17] as follows.

Definition 13: Let $R(\mathbf{U}, \mu)$ be a relational schema in T-2FRDB, r a type-2 fuzzy relation over R and ϕ a selection condition. The **selection** on r with respect to ϕ, denoted by $\sigma_\phi(r)$, is the type-2 fuzzy **relation** r' over R, including all tuples defined by

$$r' = \{t \in r \,|\, \text{Int}_{R,r,t}(\phi) \neq 0 \wedge \mu_{r'}(t) = \text{Int}_{R,r,t}(\phi)\}.$$

We note that the number 0 in $\text{Int}_{R,r,t}(\phi) \neq 0$ is also the fuzzy number 0 on $[0, 1]$.

Example 7: Consider the relation r = PATIENT in Example 4, the query "Find all patients who are *young* and diagnosed *hepatitis*" can be done by the selection operation $r' = \sigma_\phi(\text{PATIENT})$, where the selection condition ϕ = "x.AGE $\rightarrow$ *young* $\wedge$ x.DISEASE = *hepatitis*".

The selection is implemented by checking the satisfaction of all tuples in PATIENT for ϕ. From the result computed in Example 6, we can see that only the tuple t_4 satisfies ϕ with the value of membership function being $approx\_0.9$ above. Therefore, the result of the selection is the relation $r' = \sigma_\phi(\text{PATIENT})$, as shown in Table II.

TABLE II
RELATION σ_ϕ(PATIENT)

P_NAME	AGE	DISEASE	D_COST	μ
T. T. D	21	Hepatitis	30	*approx*_0.9

V. OTHER OPERATIONS ON T-2FRDB

As for the classical relational database and T-1FRDB, other basic relational algebraic operations on T-2FRDB are the projection, Cartesian product, join, intersection, union, and difference. We now extend these operations of T-1FRDB for T-2FRDB to take into account the fuzzy membership degree of tuples in relations.

1. Projection

A projection of a type-2 fuzzy relation on a set of attributes is a new type-2 fuzzy relation in which only the attributes in that set are considered for each tuple of the new relation as in the following definition.

Definition 14: Let $R = (\mathbf{U}, \mu)$ be the type-2 fuzzy relational schema, r be the relation over R and $\mathbf{L} = \{A_1, A_2, \ldots, A_k\}$ be a subset of $\mathbf{U}$. The **projection** of r on $\mathbf{L}$, denoted by $\Pi_{\mathbf{L}}(r)$, is a type-2 fuzzy relation r' over the schema R', and is determined by

1) $R' = (\mathbf{L}, \mu')$, where μ' is a mapping from $D_1 \times D_2 \times \cdots \times D_k$ to a set of all fuzzy numbers on $[0, 1]$, D_i are the value domains of A_i $(i = 1, \ldots, k)$;

2) $r' = \{t' = t[\mathbf{L}] \mid t \in r,\ \mu'_{r'}(t') = \text{MAX}_{t \in r}\{\mu_r(t) \mid t' = t[\mathbf{L}]\}\}$.

Example 8: The projection of the relation PATIENT in Table I on $\mathbf{L} = \{\text{P\_NAME}, \text{DISEASE}\}$ is the relation $\Pi_{\mathbf{L}}(\text{PATIENT})$ as shown in Table III.

TABLE III
RELATION $\Pi_{\mathbf{L}}$(PATIENT)

P_NAME	DISEASE	μ'
L. V. A	Lung cancer	0.9
L. T. B	Cirrhosis	*about*_0.5
N. T. C	Bronchitis	1.0
T. T. D	Hepatitis	high

2. Cartesian Product

For the Cartesian product of two type-2 fuzzy relations, as in the classical relational database and T-1FRDB, we assume that the sets of attributes of their schemas are disjoint as in the definition below.

Definition 15: Let $\mathbf{U}_1$, $\mathbf{U}_2$ be two sets of attributes that do not have any common element and r_1, r_2 be two fuzzy relations over two type-2 fuzzy relational schemas $R_1 = (\mathbf{U}_1, \mu_1)$ and $R_2 = (\mathbf{U}_2, \mu_2)$, respectively. The **Cartesian product** of r_1 and r_2, denoted by $r_1 \times r_2$, is a type-2 fuzzy relation r over R, and is determined by

1) $R = (\mathbf{U}, \mu)$, where $\mathbf{U} = \mathbf{U}_1 \cup \mathbf{U}_2$, μ is the mapping from $D_1 \times D_2 \times \cdots \times D_{k+m}$ to the set of all fuzzy numbers on $[0, 1]$, $k = |\boldsymbol{U}_1|$, $m = |\mathbf{U}_2|$, D_i are the value domains of $A_i \in \mathbf{U}_1 \cup \mathbf{U}_2$;

2) $r = \{t = (v_1, v_2, \ldots, v_k, v_{k+1}, v_{k+2}, \ldots, v_{k+m}) \mid t_1 = (v_1, v_2, \ldots, v_k),\ t_2 = (v_{k+1}, v_{k+2}, \ldots, v_{k+m}),\ t_1 \in r_1,\ t_2 \in r_2,\ \mu_r(t) = \text{MIN}(\mu_{1r_1}(t_1), \mu_{2r_2}(t_2))\}$.

3. Join

The join of two fuzzy relations in T-2FRDB is extended from the natural join of two relations in a classical relational database and the join in T-1FRDB. The join in T-2FRDB is defined as follows.

Definition 16: Let $\mathbf{U}_1$ and $\mathbf{U}_2$ be two sets of attributes such that value domains of two attributes of the same name A in $\mathbf{U}_1$ and $\mathbf{U}_2$, respectively, are identical. Let r_1 and r_2 be two fuzzy relations over the type-2 fuzzy relational schemas $R_1 = (\mathbf{U}_1, \mu_1)$ and $R_2 = (\mathbf{U}_2, \mu_2)$, respectively, and $\{A_k, \ldots, A_l\} = \mathbf{U}_1 \cap \mathbf{U}_2$. The natural **join** of r_1 and r_2, denoted by $r_1 \bowtie r_2$, is a type-2 fuzzy relation r over the schema R, and is determined by

1) $R = (\mathbf{U}, \mu)$, where $\mathbf{U} = \mathbf{U}_1 \cup \mathbf{U}_2$, μ is the mapping from $D_1 \times D_2 \times \cdots \times D_n$ to the set of all fuzzy numbers on $[0, 1]$, $n = |\mathbf{U}|$, D_i are the value domains of $A_i \in \mathbf{U}_1 \cup \mathbf{U}_2$;

2) $r = \{t = (v_1, \ldots, v_j, v_k, \ldots, v_l, v_m, \ldots, v_n) \mid t_1 = (v_1, \ldots, v_j, v_k, \ldots, v_l),\ t_2 = (v_k, \ldots, v_l, v_m, \ldots, v_n),\ t_1 \in r_1,\ t_2 \in r_2$ such that $v_k = t_1[A_k] = t_2[A_k], \ldots, v_l = t_1[A_l] = t_2[A_l]$ and $\mu_r(t) = \text{MIN}(\mu_{1r_1}(t_1), \mu_{2r_2}(t_2))\}$.

Example 9: Let $\mathbf{U}_1 = \{\text{P\_ID, DISEASE}\}$ and $\mathbf{U}_2 = \{\text{P\_NAME, DISEASE}\}$ be two sets of attributes, PATIENT$_1$ and PATIENT$_2$ two type-2 fuzzy relations over two schemas $R_1 = (\mathbf{U}_1, \mu_1)$ and $R_2 = (\mathbf{U}_2, \mu_2)$, as shown in Tables IV and V, respectively. It is easy to see that $\text{MIN}(\{1:1, 0.9:0.8, 0.8:0.3\}, \{0.6:0.3, 0.5:1, 0.4:0.4\}) = \{0.6:0.3, 0.5:1, 0.4:0.4\}$. So, the result of the join of PATIENT$_1$ and PATIENT$_2$ is the relation PATIENT over the schema $R = (\mathbf{U}_1 \cup \mathbf{U}_2, \mu)$ computed as in Table VI.

TABLE IV
RELATION PATIENT$_1$

P_ID	DISEASE	μ_1
PT005	Bronchitis	0.8
PT006	Gall-stone	$\{1:1, 0.9:0.8, 0.8:0.3\}$

TABLE V
RELATION PATIENT$_2$

P_NAME	DISEASE	μ_2
L. V. E	Bronchitis	0.9
N. T. F	Gall-stone	$\{0.6:0.3, 0.5:1, 0.4:0.4\}$

TABLE VI
RELATION PATIENT = PATIENT$_1$ $\bowtie$ PATIENT$_2$

P_ID	P_NAME	DISEASE	μ
PT005	L. V. E	Bronchitis	0.8
PT006	N. T. F	Gall-stone	$\{0.6:0.3, 0.5:1, 0.4:0.4\}$

4. Intersection, Union and Difference

By extending the operations of ordinary fuzzy sets in Definition 2, the set operations on the type-2 fuzzy relations in T-2FRDB are defined in turn below.

Definition 17: Let r_1 and r_2 be two type-2 fuzzy relations over the same schema $R(\mathbf{U}, \mu)$. The **intersection** of r_1 and r_2, denoted by $r_1 \cap r_2$, is a type-2 fuzzy relation r over R including tuples t's, and is defined as

$$r \cap s = \{t \mid \mu_{r\cap s}(t) = \text{MIN}(\mu_r(t), \mu_s(t))\}.$$

Definition 18: Let r_1 and r_2 be two type-2 fuzzy relations over the same schema $R(\mathbf{U}, \mu)$. The **union** of r_1 and r_2, denoted by $r \cup s$, is a type-2 fuzzy relation r over R including tuples t's, and is defined as

$$r \cup s = \{t \mid \mu_{r\cup s}(t) = \text{MAX}(\mu_r(t), \mu_s(t))\}.$$

Definition 19: Let r_1 and r_2 be two type-2 fuzzy relations over the same schema $R(\mathbf{U}, \mu)$. The **difference** of r_1 and r_2, denoted by $r - s$, is a type-2 fuzzy relation r over R including tuples t's, and is defined as

$$r - s = \{t \mid \mu_{r\cap\neg s}(t) = \text{MIN}(\mu_r(t), 1 - \mu_s(t))\}.$$

5. Properties of Algebraic Operations

In this section, we propose some properties of the fuzzy relational algebraic operations in T-2FRDB as an extension from those in the classical relational database and T-1FRDB. Clearly, these properties say that T-2FRDB model is coherent and consistent.

Theorem 1: Let r be a fuzzy relation over the schema R in T-2FRDB, 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 expressions ϕ_1 and ϕ_2 in $\phi_1 \wedge \phi_2$ are assumed to have the same tuple variable.

Proof: Let $s = \sigma_{\phi_2}(r)$. We have

$$\begin{aligned}
&\sigma_{\phi_1}(\sigma_{\phi_2}(r)) \\
&\quad = \{t \in s | \operatorname{Int}_{R,s,t}(\phi_1) \neq 0\} \qquad \text{(by Definition 13)} \\
&\quad = \{t \in r | \operatorname{Int}_{R,r,t}(\phi_2) \neq 0 \wedge \operatorname{Int}_{R,s,t}(\phi_1) \neq 0\} \\
&\quad = \{t \in r | \operatorname{Int}_{R,r,t}(\phi_2) \neq 0 \wedge \operatorname{Int}_{R,r,t}(\phi_1) \neq 0)\} \\
&\qquad\qquad \text{(because } s \subseteq r) \\
&\quad = \{t \in r | \text{MIN}(\operatorname{Int}_{R,r,t}(\phi_2), Int_{R,r,t}(\phi_1)) \neq 0)\} \\
&\qquad\qquad \text{(by Definition 12)} \\
&\quad = \{t \in r | \operatorname{Int}_{R,r,t}(\phi_2 \wedge \phi_1) \neq 0)\} = \sigma_{\phi_1\wedge\phi_2}(r).
\end{aligned}$$

So, $\sigma_{\phi_1}(\sigma_{\phi_2}(r)) = \sigma_{\phi_1\wedge\phi_2}(r)$. Similarly, we have $\sigma_{\phi_2}(\sigma_{\phi_1}(r)) = \sigma_{\phi_2\wedge\phi_1}(r)$. Since $\phi_1 \wedge \phi_2$ if and only if $\phi_2 \wedge \phi_1$ (*i.e.*, the logical conjunction of selection conditions is commutative), we have $\sigma_{\phi_1\wedge\phi_2}(r) = \sigma_{\phi_2\wedge\phi_1}(r)$.

Therefore, it results in $\sigma_{\phi_1}(\sigma_{\phi_2}(r)) = \sigma_{\phi_2}(\sigma_{\phi_1}(r))$ and so $\sigma_{\phi_1}(\sigma_{\phi_2}(r)) = \sigma_{\phi_2}(\sigma_{\phi_1}(r)) = \sigma_{\phi_1 \wedge \phi_2}(r)$. □

Theorem 2: Let r be a fuzzy relation over the schema R in T-2FRDB, $\mathbf{A}$ and $\mathbf{B}$ be two subsets of attributes of R, and $\mathbf{A} \subseteq \mathbf{B}$. Then

$$\Pi_{\mathbf{A}}(\Pi_{\mathbf{B}}(r)) = \Pi_{\mathbf{A}}(r). \tag{2}$$

Proof: Because $\mathbf{A} \subseteq \mathbf{B}$, so $\mathbf{A} \cap \mathbf{B} = \mathbf{A}$. By Definition 14, it is easy to see that both sides of (2) are relations over the same schema with the set of attributes $\mathbf{A} \cap \mathbf{B} = \mathbf{A}$. By the property of the projection of classical relations (Definitions 9 and 14), it follows that two classical sets of tuples which are collected from two relations $\Pi_{\mathbf{A}}(\Pi_{\mathbf{B}}(r))$ and $\Pi_{\mathbf{A}}(r)$ are the same. Also by Definition 14, the operation MAX in both sides of (2) is executed on the same value set of the membership degrees of tuples of r. Therefore, $\Pi_{\mathbf{A}}(\Pi_{\mathbf{B}}(r)) = \Pi_{\mathbf{A} \cap \mathbf{B}}(r) = \Pi_{\mathbf{A}}(r)$. □

Theorem 3: Let R_1, R_2 and R_3 be the schemas in T-2FRDB such that if they have attributes of the same name, such attributes have the same value domain. Let r_1, r_2 and r_3 be the fuzzy relations over R_1, R_2 and R_3 respectively. Then

$$r_1 \bowtie r_2 = r_2 \bowtie r_1, \tag{3}$$

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

Proof: Clearly, $r_1 \bowtie r_2$ and $r_2 \bowtie r_1$ are two relations over the same schema. By the property of the join of classical relations (Definitions 9 and 16), it follows that two classical sets of tuples which are collected from two relations $r_1 \bowtie r_2$ and $r_2 \bowtie r_1$ are the same. In addition, the operation MIN of two fuzzy numbers (two membership degrees of two tuples in r_1 and r_2, respectively) has commutativity. From that the join of tuples has commutativity. So, by Definition 16 we have $r_1 \bowtie r_2 = r_2 \bowtie r_1$.

By Definition 16, clearly $(r_1 \bowtie r_2) \bowtie r_3$ and $r_1 \bowtie (r_2 \bowtie r_3)$ are two relations over the same schema. By the property of the join of classical relations (Definitions 9 and 16), it follows that two classical sets of tuples which are collected from two relations $(r_1 \bowtie r_2) \bowtie r_3$ and $r_1 \bowtie (r_2 \bowtie r_3)$ are the same. Let A be a common attribute in $\boldsymbol{U}_1$, $\boldsymbol{U}_2$ and $\boldsymbol{U}_3$ of R_1, R_2 and R_3. Because the operation MIN of two fuzzy numbers and the identical operation of attribute values have associativity, the join of tuples has associativity. Thus, by Definition 16, we have $(r_1 \bowtie r_2) \bowtie r_3 = r_1 \bowtie (r_2 \bowtie r_3)$. □

Because the Cartesian product (Definition 15) is a particular case of the join, we have the straight corollary of Theorem 3 as follows.

Corollary 1: Let R_1, R_2 and R_3 be schemas in T-2FRDB such that each pair of them does not have any common attribute, r_1, r_2 and r_3 be fuzzy 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 relations over the same schema R in T-2FRDB. Then

$$r_1 \cap r_2 = r_2 \cap r_1, \tag{7}$$

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

$$r_1 \cup r_2 = r_2 \cup r_1, \tag{9}$$

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

Proof: Because the intersection and union operations of sets and the MIN and MAX operations of fuzzy numbers have commutativity and associativity. So, by Definitions 17 and 18, Equations (7), (8), (9) and (10) then follow. □

VI. Conclusions

In this paper, we have proposed a type-2 fuzzy relational database model, abbreviated by T-2FRDB, as an extension of the type-1 fuzzy relational database models. In T-2FRDB, the membership degrees of tuples in a relation are represented by the fuzzy numbers on the interval $[0, 1]$. The data model and fuzzy relational algebraic operations in T-2FRDB have been defined formally and consistently. Computing and associating the membership degrees of tuples in manipulating of the algebraic operations are implemented by the operations MAX and MIN using the extension principle. T-2FRDB allows us to express and execute the soft queries that are associated with fuzzy sets for dealing with imprecise information in real databases. A set of basic properties of the algebraic operations in T-2FRDB has also been proposed as theorems, which have been completely proven.

In subsequent studies, we will extend notions of the key and fuzzy functional dependencies in T-1FRDB for T-2FRDB, and build a type-2 fuzzy relational database management system based on T-2FRDB with the familiar querying and manipulating language like SQL for representing and handling the imprecise information in real world applications.

References

[1] C. J. Date, *An introduction to database systems*, 8th ed. Addison Wesley, 2003.
[2] 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.
[3] G. J. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*. Prentice-Hall, Inc., 1994.
[4] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[5] S. Chakraborty, "Codd's relational data model and fuzzy logic: A practical approach to find the computer solution," *International Journal of Advanced Technology & Engineering Research (IJATER)*, vol. 2, no. 4, pp. 21–27, 2012.
[6] J. C. Cubero, J. M. Medina, O. Pons, and M.-A. Vila, "Data summarization in relational databases through fuzzy dependencies," *Information Sciences*, vol. 121, no. 3, pp. 233–270, 1999.
[7] D. Dubois and H. Prade, "Using fuzzy sets in flexible querying: Why and how?" in *Flexible query answering systems*. Springer, 1997, pp. 45–60.
[8] J. Mishra and S. Ghosh, "Uncertain query processing using vague set or fuzzy set: which one is better?" *International Journal of Computers Communications & Control*, vol. 9, no. 6, pp. 730–740, 2014.
[9] F. E. Petry, "Fuzzy databases: Principles and applications," *International series in intelligent technologies*, 1996.
[10] K. Raju and A. K. Majumdar, "Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems," *ACM Transactions on Database Systems (TODS)*, vol. 13, no. 2, pp. 129–166, 1988.
[11] P. Saxena and D. K. Tayal, "Normalization in type-2 fuzzy relational data model based on fuzzy functional dependency using fuzzy functions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 20, no. 01, pp. 99–138, 2012.
[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, pp. 2037–2044, 2010.
[13] O. Pivert and H. Prade, "Dealing with aggregate queries in an uncertain database model based on possibilistic certainty," in *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer, 2014, pp. 150–159.
[14] 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.
[15] S. M. Darwish, T. F. Mabrouk, and Y. F. Mokhtar, "Enriching Vague Queries by Type-2 Fuzzy Orderings," *Lecture Notes on Information Theory Vol*, vol. 2, no. 2, 2014.
[16] A. Niewiadomski, "A type-2 fuzzy approach to linguistic summarization of data," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 1, pp. 198–212, 2008.
[17] N. Ha, "A fuzzy relational database model," *Journal of Information and Communications Technology*, vol. 5, no. 1, pp. 37–45, 2015.
[18] S. Parsons, "Current approaches to handling imperfect information in data and knowledge bases," *IEEE Transactions on knowledge and data engineering*, vol. 8, no. 3, pp. 353–372, 1996.
[19] A. Konar, *Computational intelligence: principles, techniques and applications*. Springer, 2005.
[20] T. J. Ross, *Fuzzy logic with engineering applications*, 3rd ed. John Wiley & Sons, 2010.
[21] N. N. Karnik and J. M. Mendel, "Operations on type-2 fuzzy sets," *Fuzzy sets and systems*, vol. 122, no. 2, pp. 327–348, 2001.

Nguyen Hoa is a lecturer at Saigon University. He received the B.Sc. degree in Mathematics in 1982 from Vinh Pedagogical University, the M.Eng. degree in Computer Science in 2003 from Ho Chi Minh City University of Technology, and the Ph.D. degree in Computer Science in 2008 from Vietnam National University, Ho Chi Minh City. His research interests include imprecise and uncertain knowledge representation, fuzzy databases and probabilistic databases.