MTA SZTAKI Közlemények 36/1987 pp. 197-222

AN EXTENDED RELATIONAL DATABASE BY APPLICATION OF FUZZY SET THEORY AND LINGUISTIC VARIABLE

*LIE TIEN VUONG** and *HO THUAN***

*Institute of Computer Science and
Cybernetic - Hanoi - Vietnam

**Computer and Automation Institute
Hungarian Academy of Sciences

1. INTRODUCTION

The relational database have been studied since Codd [4]. Such database can only deal with well-defined and unambiguous data. But in the real world there exist data which can not be defined in certain and well-defined form by any means. The databases for above mentioned data have been investigated by different authors. [8,15] have developed the models for data with incomplete information and null-values. In [10,13,14] the authors have used the concept of linguistic variables to design intelligent database systems. The use of linguistic variable for a database is complicated but it is every important for describing objects that we do not have enough information such as "he is young", "A is far from B" ... These objects may be presented in a table as below:

STUDENT	NAME	AGE	HEIGHT
	A	20	about 1,70 m
	B	young	1,80
	C	about 25	high

The terms "young", "height", "about 25",... are called fuzzy terms. The fuzzy terms are a great class of data. To extened a database with fuzzy terms, the authors use in this paper the possibility distribution function [16] and multivalued

This work was supported by the National Scientific Research Fund No. 1066/1985

logic.

In section 2, the basic definitions of fuzzy set theory and linguistic variables are briefly mentioned. In section 3, we introduce the conceptual framework for a fuzzy database. The evaluation of a fuzzy query in a fuzzy database by relational algebra is presented in section 4. The section 5 extends the concept of data dependencies in relational database. In this section a concept of ternary degenerate decomposition of an extended relation is also introduced.

2. THE BASIC DEFINITION OF FUZZY SETS

In this section we shall briefly present the fuzzy notations and concepts which are minimally required for this paper. More details of discussions may be seen in [9,16].

Definition 2.1.

Let $U = \{u\}$ be a universe of discourse. A fuzzy set $\underset{\sim}{u}$ of U is a set of ordered pairs $\{(u, \mu_{\underset{\sim}{u}}(u))\}$, $u \in U$, where $\mu_{\underset{\sim}{u}}(u)$ is the grade of membership of u in $\underset{\sim}{u}$, and $\mu_{\underset{\sim}{u}} : U \to [0,1]$ is the membership function.

Definition 2.2.

Let $\underset{\sim}{u}$ and $\underset{\sim}{v}$ be two fuzzy sets of U.

a. Equality:

$\underset{\sim}{u}$ and $\underset{\sim}{v}$ are equal, written as $\underset{\sim}{u} \overset{f}{=} \underset{\sim}{v}$, iff $\mu_{\underset{\sim}{u}}(u) = \mu_{\underset{\sim}{v}}(u)$, $\forall u \in U$.

b. Containment:

$\underset{\sim}{u}$ is a subset of $\underset{\sim}{v}$, written as $\underset{\sim}{u} \overset{f}{\subseteq} \underset{\sim}{v}$, iff $\mu_{\underset{\sim}{u}}(u) \leq \mu_{\underset{\sim}{v}}(u)$ $\forall u \in U$.

c. Complementation:

The complement of a fuzzy set $\underset{\sim}{u}$ of U, denoted by $\overset{f}{\neg}\underset{\sim}{u}$, is defined by $\mu_{\overset{f}{\neg}\underset{\sim}{u}}(u) = 1 - \mu_{\underset{\sim}{u}}(u)$, $\forall u \in U$.

d. Union:

The union of $\underset{\sim}{u}$ and $\underset{\sim}{v}$, denoted by $\underset{\sim}{u} \overset{f}{\cup} \underset{\sim}{v}$, is defined by $\mu_{\underset{\sim}{u} \overset{f}{\cup} \underset{\sim}{v}}(u) = \mu_{\underset{\sim}{u}}(u) \vee \mu_{\underset{\sim}{v}}(u)$, $\forall u \in U$.

e. Intersection:

The intersection of $\underset{\sim}{u}$ and $\underset{\sim}{v}$, denoted by $\underset{\sim}{u} \overset{f}{\cap} \underset{\sim}{v}$, is defined by $\mu_{\underset{\sim}{u} \overset{f}{\cap} \underset{\sim}{v}}(u) = \mu_{\underset{\sim}{u}}(u) \wedge \mu_{\underset{\sim}{v}}(u)$, $u \in U$.

The symbols $\vee$ and $\wedge$ denote the maximum and the minimum, respectively.

Definition 2.3.

Let $\underset{\sim}{u}$ be a given fuzzy set of U. A λ-level fuzzy set, denoted by $\underset{\sim}{u}_\lambda$, is defined by

$$\underset{\sim}{u}_\lambda = \{(u, \mu_{\underset{\sim}{u}}(u)) \mid u \in u(\lambda)\}$$

where λ-level set $u(\lambda)$ is defined by

$$u(\lambda) = \{u \mid \mu_{\underset{\sim}{u}}(u) \geq \lambda, \quad u \in U\}, \qquad \lambda \in [0,1].$$

Definition 2.4

A linguistic variable is characterized by a quintuple (A,T(A),U,G,M) in which A is the name of the variable; T(A) (or simply T) denotes the term set of A, that is, the set of names of linguistic values of A; G is a syntactic rule for generating the names in T; M is a semantic rule for associating, with each t in T, its meaning M(t), which is a fuzzy set of U.

The meaning of a fuzzy term can be presented in the form $M(t) = \{(u, \mu_t(u)) \mid u \in U\}$.

It is easy to express a λ-level meaning of a fuzzy term $t \in T$. Let t be a linguistic value of A of universe discourse U. The λ-level meaning (or simply λ-meaning) $M_\lambda(t)$, $t \in T$ is a fuzzy set in the form $M_\lambda(t) = \{(u, \mu_t(u))\ u \in M_t(\lambda)\}$, where $M_t(\lambda)$ denoting λ-level set of fuzzy term t, is defined by

$$M_t(\lambda) = \{u \mid \mu_t(u) \geq \lambda,\ u \in U\}.$$

3. AN EXTENDED DATABASE BY APPLICATION OF FUZZY SETS AND LINGUISTIC VARIABLES

A relation over a set of attributes $W = \{A_1, \ldots, A_n\}$ is denoted by $R(W)$ (or simply R). Each attribute $A \in W$ is associated with a basic domain $U(A)$ which specifies all possible real values for A. Each basic domain can be extended by a corresponding set of linguistic values $T(A)$. Then the domain of the attribute A can be presented in the form $Dom(A) = U(A) \cup T(A)$ (or simply $D = U \cup T$).

A relation R over a set of attributes W is said to be a full relation of it contains no linguistic values, that is, for any $A \in W$, $Dom(A) = U(A)$ (i.e. $T(A) = \emptyset$). If R is not a full relation, that is $T(A) \neq \emptyset$ for some $A \in W$, then R is called extended relation.

We use A,B,C,... (or with indexes) to denote single attribute and X,Y,Z,.. (or with indexes) to denote sets of attributes of W. For a set of attributes $X \subseteq W$, a X-value is a mapping r that assigns to each attribute A_i of X an value from its domain $D(A_i) = U(A_i) \cup T(A_i)$ (or simply $D_i = U_i \cup T_i$). The value assigned to the attribute by such a mapping is denoted by $r[A_i]$. An extended relation over X is a set of X-values.

Without loss of generality it is assumes that the set of attributes W is finite. An extended relational database can be defined as follows:

Definition 3.1.

An extended relational database DB is defined as a set of extended relations R_i, $i = \overline{1,n}$, i.e.

$$DB = \{R_1, \ldots, R_n\},$$

in which every relation R_i is defined as a subset of the Cartesian product of a collection of domains, i.e.

$$R_i \subseteq \{U(A_{i_1}) \cup T(A_{i_1})\} \times \ldots \times \{U(A_{i_k}) \cup T(A_{i_k})\},$$

where $U(A_{i_j})$, $j = \overline{1,k}$ are basic domains and $T(A_{i_j})$, $j = \overline{1,k}$ are the set of fuzzy terms (linguistic values) of linguistic variables A_{i_j}.

To evaluate the meaning of any fuzzy term $t \in T(A_i)$, $A_i \in X$, in this paper can be used the techniques developed by [16]. The meaning of all values $u \in U(A_i)$, $A_i \in X$, is denoted by $M(u)$ and presented in the special form $M(u) = \{(u,1)\}$.

We introduce some (mathematical) concepts as follows.

Definition 3.2.

Let r_1, r_2 be two tuples of an extended relation $R(X)$ over the set of attributes $X \subseteq W$.

a. $r_1[A] \overset{\lambda}{\approx} r_2[A]$ iff $M_\lambda(r_1[A]) = M_\lambda(r_2[A])$ for $A \in X$,

$r_1[A]$, $r_2[A] \in D(A)$, $\lambda \in [0,1]$.

b. $r_1, r_2 \in R(X)$, $r_1 \overset{\lambda}{\approx} r_2$ iff $r_1[A] \overset{\lambda}{\approx} r_2[A]$ for all $A \in X$.

The relation $\overset{\lambda}{\approx}$ is called λ-level equivalence (or briefly λ -equivalence).

Remark.

If the relation R is full, then the concept of λ-level equivalence is identified with the equality of two real values, i.e. $r_1[A] = r_2[A]$, $r_1[A]$, $r_2[A] \in U(A)$.

It is easy to show that the relation $\overset{\lambda}{\approx}$ is an equivalence relation.

In the following $\overset{\lambda}{\approx}$ is written by $\approx$ for sake of simplicity. Let W^* be the set of all possible tuples which are defined on W, i.e. it contains all X-values for all $X \subseteq W$. Every λ-level extended (written x-relation for short) is a λ-equivalence class defined by $\approx$. The class of relations equivalent to R is denoted $\bar{R}$ and R is called a representation of $\bar{R}$. Two relations R_1, R_2 over X are λ-equivalent, denoted by $R_1 \approx R_2$, iff

$$\text{for } \forall r_1 \in R_1,\ \exists r_2 \in R_2 \quad \text{such that} \quad r_1 \approx r_2 \quad \text{and}$$
$$\text{for } \forall r_2 \in R_2,\ \exists r_1 \in R_1 \quad \text{such that} \quad r_1 \approx r_2.$$

Given a set of tuples $\{r_1,\dots,r_n\}$, one can eliminate all tuples that are λ-equivalent to other tuples, and enlarge the others to their λ-equivalent X-values. The x-relation represented by the set of X-values so obtained will be denoted $\{r_1,\dots,r_n\}_f$.

A tuple t is said to belong to or to be an element of $\bar{R}$, written $t \in \bar{R}$ when for some R' in $\bar{R}$, $t \in R'$

The following proposition is straightforward.

Proposition 3.1.

t is a tuple of $\bar{R}$ iff there exists a tuple r of R such that $r \approx t$.

In other words, a tuple t belongs to an x-relation iff its representation contains a tuple r which is λ-equivalent to t. An x-relation over $X \subseteq W$ is represented by the set of

X-values and will be denoted by the set $\{r_1,\dots,r_n\}_f$, in which all λ-equivalent tuples have been identified. The set operations on the set of x-relations can be defined as follows.

Let $\bar{R}_1$, $\bar{R}_2$ be two x-relations over X. We have

Union: $\bar{R}_1 \cup \bar{R}_2 = \{r \mid r \in R_1 \text{ or } r \in R_2\}_f$.

Intersection: $\bar{R}_1 \cap \bar{R}_2 = \{r \mid \exists r_1 \in R_1,\ r \approx r_1 \text{ and } \exists r_2 \in R_2,\ r \approx r_2\}_f$

Diference: $\bar{R}_1 \setminus \bar{R}_2 = \{r \mid r \in R_1 \text{ and } \nexists r_2 \in R_2 \text{ such that } r \approx r_2\}_f$.

Given a set of all λ-meanings of a linguistic variable A (domain of A is $D = U \cup T$), denoted by $\mathcal{D}_\lambda$. Let $M_\lambda(u)$ and $M_\lambda(v)$ be λ-meanings of $u, v \in D$, respectively. Some operations in $\mathcal{D}_\lambda$ can be defined as follows:

Definition 3.3.

Union:

$$M_\lambda(u) \overset{f}{\cup} M_\lambda(v) = \{(u_i, \mu_u(u_i) \vee \mu_v(u_i)) \mid u_i \in M_u(\lambda) \cup M_v(\lambda)\}.$$

Intersection:

$$M_\lambda(u) \overset{f}{\cap} M_\lambda(v) = \{(u_i, \mu_u(u_i) \wedge \mu_v(u_i)) \mid u_i \in M_u(\lambda) \cap M_v(\lambda)\}$$

Complementation:

$$\overset{f}{\neg} M_\lambda(u) = \{(u_i,\ 1-\mu_u(u_i)) \mid u_i \in U,\ 1-\mu_u(u_i) \geq \lambda\}.$$

It is assumed that there exist in $\mathcal{D}_\lambda$ two elements MO and M1 of two values u_o and $u_1 \in D$, where MO and M1 are defined by:

$$MO = M(u_o) = \emptyset.$$

$$M1 = M(u_1) = \{(u_i, \mu_{u_1}(u_i)) \mid \mu_{u_1}(u_i) = 1 \text{ for } \forall u_i \in U\}.$$

MO is called λ-empty meaning and M1 is called λ-full meaning. Clearly, for all $M(u) \in \mathcal{D}_\lambda$ (the set of all λ-meanings of a linguistic variable A)

$$M(u) = M0 \text{ iff } \mu_u(u_i) < \lambda \text{ for } \forall u_i \in U \quad \text{and}$$

$$M(u) = M1 \text{ iff } \mu_u(u_i) = 1 \text{ for } \forall u_i \in U.$$

The λ-empty meaning and λ-full meaning have following properties:

For all $M(u) \in \mathcal{D}_\lambda$

$$M(u) \overset{f}{\cap} M0 = M0; \; M(u) \overset{f}{\cup} M0 = M(u);$$

$$M(u) \overset{f}{\cap} M1 = M(u); \; M(u) \overset{f}{\cup} M1 = M1.$$

Proposition 3.2.

The set of all λ-meanings $\mathcal{D}_\lambda$ of a linguistic variable A of an x-relation $\bar{R}(X)$ with the operations $\overset{f}{\cup}$, $\overset{f}{\cap}$ is a distributive lattice but not a Boolean algebra with the operations $\overset{f}{\cup}$, $\overset{f}{\cap}$ and $\overset{f}{\neg}$.

Proof.

It is easy to show, that all laws such as idempotency, commutativity, associativity, absorption and distributivity for the operations $\overset{f}{\cup}$ and $\overset{f}{\cap}$ are satisfied.

In order to show that $\mathcal{D}_\lambda$ is not Boolean algebra, we consider a following example.

Let us assume that there exist two elements of λ-empty meaning M0 and λ-full meaning M1. We must show that the laws of complementarity are not satisfied, i.e. $M(u) \overset{f}{\cap} \overset{f}{\neg} M(u) \neq M0$, and $M(u) \overset{f}{\cup} \overset{f}{\neg} M(u) \neq M1$ for some $M(u) \in \mathcal{D}_\lambda$.

Let $\lambda = 0.5$,

$$M(u) = \{(u_1,0.3),(u_2,0.6),(u_3,0.7),(u_4,0.8),(u_5,1.),(u_6,0.5),(u_7,0.4)\}$$

$$M_{0.5}(u) = \{(u_2,0.6),(u_3,0.7),(u_4,0.8),(u_5,1.),(u_6,0.5)\}$$

$$\overset{f}{\neg} M_{0.5}(u) = (u_1,0.7),(u_6,0.5),(u_7,0.6)$$

$$M_{0.5}(u) \overset{f}{\cap} \overset{f}{\neg} M_{0.5}(u) = \{(u_6,0.5)\} \neq M0.$$

$$M_{0.5}(u) \overset{f}{\cup} \overset{f}{\neg} M_{0.5}(u) = \{(u_1,0.7),(u_2,0.6),(u_3,0.7),(u_4,0.8),$$
$$(u_5,1.),(u_6,0.5),(u_7,0.6)\} \neq M1.$$

4. QUERY EVALUATION

If a query Q with fuzzy terms is formulated on an extended database then there are three important bounds of interest:

(a). The set of all objects which surely satisfy the query Q, i.e. they satisfy Q with truth-value 1.

(b). The set of all objects which very probably satisfy the query Q, i.e. they satisfy Q with truth-value equal or greater than λ $(0 < \lambda < 1)$.

(c). The set of all objects which may possibly satisfy the query Q, i.e. they satisfy Q with truth-value less than λ.

In this paper we are interested only in the problems (a) and (b) for a language based upon the relational algebra. The λ-value depends on database users. The problem (c) is very complicated and its part is investigated together with null--value problem in [8,14,15].

To evaluate a query in an extended database, the operations ∇ and Δ must be used here for defining the upper element and lower element of any two elements of lattice $\mathcal{D}_\lambda$ ($\mathcal{D}_\lambda$ is a partially ordered set), respectively (see [7]).

Let $M_\lambda(u)$, $M_\lambda(v)$ be two λ-meanings in $\mathcal{D}_\lambda$ in the form

$$M_\lambda(u) = \{(u_i,\ \mu_u(u_i)) \mid u_i \in M_u(\lambda)\}, \quad u \in D \quad \text{and}$$

$$M_\lambda(v) = \{(v_j,\ \mu_v(v_j)) \mid v_j \in M_v(\lambda)\} \quad v \in D$$

or in other form (see definition 2.2)

$$M_\lambda(u) = \overset{f}{\bigcup_{u_i \in M_u(\lambda)}} \{(u_i, \mu_u(u_i))\} \quad \text{and}$$

$$M_\lambda(v) = \overset{f}{\bigcup_{v_j \in M_v(\lambda)}} \{(v_j, \mu_v(v_j))\}.$$

The operations are defined as follows:

$$M_\lambda(u) \nabla M_\lambda(v) = \overset{f}{\bigcup_{u_i \in M_u(\lambda)}} \{(u_i \vee v_j,\ \mu_u(u_i) \wedge \mu_v(v_j)) \mid v_j \in M_v(\lambda)\} \quad \text{and}$$

$$M_\lambda(u) \Delta M_\lambda(v) = \overset{f}{\bigcup_{u_i \in M_u(\lambda)}} \{(u_i \wedge v_j,\ \mu_u(u_i) \wedge \mu_v(v_j)) \mid v_j \in M_v(\lambda)\},$$

for all $u, v \in D$.

The predicate calculus based on languages contains two simple relational expressions such as $r_1[A]\Theta\ r_2[B]$ and $r[A]\Theta c$, where r_1, r_2 and r are tuple variables, A and B are attributes, c is a fuzzy constant from domain D(A), Θ is one of the comparision operations $=$, $\neq$, $>$, $\geq$, $<$, $\leq$. The evaluation value of an expression is in the interval [0,1].

W.l.o.g. the above expressions can be presented in the form of a simple fuzzy predicate, denoted by p:

$$p =: u\ \Theta\ v \quad \text{or} \quad p := u\ \Theta\ c, \quad \text{where} \quad u, v, c \in D.$$

Let f be an evaluation function with respect to p. The truth-

-value of function f with respect to p is a number τ of the set $\{0\} \cup [\lambda, 1]$. The Boolean operators AND, OR, NOT may be defined as τ AND $\tau' = \tau \wedge \tau'$, τ OR $\tau' = \tau \vee \tau'$ and

$$\mathrm{NOT}\tau = \begin{cases} 1-\tau & \text{if } 1-\tau \geq \lambda \\ 0, & \text{otherwise} \end{cases}$$

where τ, τ' are numbers of the set $\{0\} \cup [\lambda, 1]$.

Now we consider the comparison operations $\Theta \in \{=, \neq, >, \geq, <, \leq\}$. These are defined as follows.

$$f(u = v) = \begin{cases} 1, & \text{if } M_\lambda(u) = M_\lambda(v),\ u, v \in D \\ \tau, & \text{if } \quad 1 > \tau \geq \lambda \\ 0, & \text{otherwise} \end{cases}$$

where τ is defined by

$$\tau = \frac{1}{1 + \left(\sum_{u_i \in M_u(\lambda) \cup M_v(\lambda)} (\mu_u(u_i) - \mu_v(u_i))^2 \right)^{1/2}}$$

$$f(u > v) = \begin{cases} 1, & \text{if } M_\lambda(u) \Delta M_\lambda(v) = M_\lambda(v) \quad \text{and} \quad M_\lambda(u) \neq M_\lambda(v) \\ 0, & \text{otherwise.} \end{cases}$$

$f(u < v)$ is defined analogously.

$$f(u \neq v) = \begin{cases} 0, & \text{if } M_\lambda(u) = M_\lambda(v), \\ \tau' = 1-\tau & \text{if } 1-\tau \geq \lambda \\ 1, & \text{otherwise.} \end{cases}$$

The remaining operations can be defined by

$$f(u \leq v) = f(u = v) \vee f(u < v),$$

$$f(u \geq v) = f(u = v) \vee f(u > v).$$

Based upon evaluation of a simple fuzzy predicate we can evaluate a fuzzy predicate expression as follows:

Let p and q be two simple fuzzy predicates. Then we have

$$f(p \text{ AND } q) = f(p) \wedge f(q).$$

$$f(p \text{ AND } q) = f(p) \vee f(q).$$

Given two tuples of W^* $r_1[X]$ and $r_2[X]$ W^* with $X \subseteq W$. The truth-value of an expression $r_1[X]\Theta r_2[X]$ is defined by the following equality

$$f(r_1[X]\Theta r_2[X]) = \bigwedge_{A_i \in X} (f(r_1[A_i]\Theta r_2[A_i])).$$

where $\Theta \in \{=, \neq, >, \geq, <, \leq\}$.

From the above concepts of an expression evaluation we can define the following relational operations.

Definition 4.1.

Let $\bar{R}(Y)$ be an x-relation. Let X be a subset of $Y \subseteq W$. The projection of $\bar{R}(Y)$ on X, denoted by $\bar{R}[X]$, is a set of tuples r for which

- there exists r' in $\bar{R}(Y)$ such that $r[X] \overset{\lambda}{\approx} r'[X]$,

- there exists no r'' in $\bar{R}(Y)$ such that $r''[X] \neq r'[X]$, $r''[X] \overset{\lambda}{\approx} r[X]$

i.e.

$$\bar{R}[X] = \{r[X] = (r[A_1],\dots,r[A_k]) \mid r \in R \text{ and } A_i \in X,\ i=\overline{1,k}\}_f.$$

Definition 4.2.

Let $\bar{R}$ be an x-relation over X. A,B are two attributes of X and c is a fuzzy constant of D(A). The selection $A \Theta B$ and $A \Theta c$ can be defined by

$$\bar{R}[A\Theta B] = \{r \mid f(r[A]\Theta r[B]) \geq \tau, \quad r \in R, \quad \tau \geq \lambda\}_f ,$$

$$\bar{R}[A\Theta c] = \{r \mid f(r[A]\Theta c) \geq \tau, \quad r \in R, \quad \tau \geq \lambda\}_f ,$$

respectively, where Θ is one of $\{=, \neq, >, \geq, <, \leq\}$.

Definition 4.3.

Let $\bar{R}$ and $\bar{S}$ be two x-relations over XY[1] and YZ, respectively, where X,Y and Z are subsets of W. The natural join of x-relation $\bar{R}$ and x-relation $\bar{S}$ on the common set of attributes is defined by

$$\begin{aligned} R[XY]*S[YZ] = \{r \mid &(\exists t \in R)(\exists s \in S)[(\forall A \in Y)(r[A] = t[A] \text{ or } \\ &r[A] = s[A])(f(t[A] = s[A]) = 1)] \text{ and } \\ &(\forall A \in X)[r[A] = t[A]] \text{ and } \\ &(\forall A \in Z)[r[A] = s[A]]\}_f . \end{aligned}$$

5. THE DATA DEPENDENCIES IN AN EXTENDED RELATIONAL DATABASE

5.1. Lossless decomposition of an extended relation

The concept of loss-less decomposition of a relation is very important in the process for designing a database, because instead of storing the relation R in the database, we can store only its projections.

---

1) *The union* $X \cup Y$ *is denoted by* XY.

In this paper we only investigate the lossless decomposition of an x-relation into a family of some of its projections. Let an x-relation $\bar{R}$ be a representation of a λ-equivalence class under $\approx$ in the universe x-relation $R(W)$. $r_1[A] \stackrel{\lambda}{=} r_2[A]$ means that $r_1[A] \approx r_2[A]$, i.e. $M_\lambda(r_1[A]) = M_\lambda(r_2[A])$ for $r_1, r_2 \in R$ and $A \in W$. When no confusion occurs, in this section we write the symbol R instead of $\bar{R}$, being a λ-level extended relation.

Definition 5.1.

Let X,Y be two subsets of W with $XY = W$. An x-relation $R(W)$ is said lossless decomposable (or simply: decomposable), denoted by $R(W) = R[X]*R[Y]$, if for all tuples pairs $r_1, r_2 \in R$ satisfying $r_1[X \cap Y] \stackrel{\lambda}{=} r_2[X \cap Y]$, there is a tuple $r \in R$ such that $r[X] \stackrel{\lambda}{=} r_1[X]$ and $r[Y] \stackrel{\lambda}{=} r_2[Y]$.

From the above definition 5.1 the concept of decomposition of an x-relation into n different projections can be generalized as follows.

Definition 5.2.

Let X_i, $i = \overline{1,n}$ be subsets of W with $\bigcup_{i=1}^{n} X_i = W$. An x-relation $R(W)$ is said n-ary decomposable if for any n tuples $r_i \in R$, $i = \overline{1,n}$ such that $r_i[X_i \cap X_j] \stackrel{\lambda}{=} r_j[X_i \cap X_j]$, $i \neq j$, $i,j = \overline{1,n}$, there is a tuple $r \in R$ such that $r[X_i] \stackrel{\lambda}{=} r_i[X_i]$, $i = \overline{1,n}$.

Some important data dependencies can be defined as follows:

Definition 5.3.

Let $X \subseteq W$, $Y \subseteq W$. A λ-functional dependency (abbr. λFD) $X \overset{\lambda}{\rightarrow} Y$ holds in an x-relation $R(W)$ if for every two tuples $r_1, r_2 \in R$, $r_1[X] \overset{\lambda}{=} r_2[X]$ implies $r_1[Y] \overset{\lambda}{=} r_2[Y]$.

Definition 5.4.

Let $X \subseteq W$, $Y \subseteq W$ and $Z = W \setminus XY$.
A λ-multivalued dependency (abbr. λMVD) $X \overset{\lambda}{\twoheadrightarrow} Y|Z$ holds in an x-relation $R(W)$ if for every pair of tuples

$r_1 = (r_1[X], r_1[Y], r_1[Z])$ and $r_2 = (r_2[X], r_2[Y], r_2[Z])$ belong to R such that $r_1[X] \overset{\lambda}{=} r_2[X]$,

$r_3 = (r_1[X], r_1[Y], r_2[Z])$ and $r_4 = (r_1[X], r_2[Y], r_1[Z])$ belong to R, too.

A different way to view an λMVD and a decomposition is given below, which again is a generalization of a similar result of R. Fagin.

Proposition 5.1.

λMVD $X \overset{\lambda}{\twoheadrightarrow} Y$ holds in an x-relation $R(W)$, where $X \subseteq W$, $Y \subseteq W$, if and only if, whenever R is lossless decomposable in two projections $R[XY]$ and $R[X(W \setminus XY)]$.

Proof.

It is assumed that R is decomposable.
From definition 5.1 there are two tuples $r_1, r_2 \in R$ such that $r_1 = (r_1[X], r_1[Y], r_1[Z])$ and $r_2 = (r_2[X], r_2[Y], r_2[Z])$, where $Z = W \setminus XY$, $r_1[X] \overset{\lambda}{=} r_2[X]$. Then there is tuple $r_3 \in R$ such that

$r_3[XY] \overset{\lambda}{=} r_1[XY] = (r_1[X], r_1[Y])$ and

$r_3[XZ] \overset{\lambda}{=} r_{.}[XZ] = (r_1[X], r_2[Z])$.

This means, that $r_3 = (r_1[X], r_1[Y], r_2[Z]) \in R$.
Similarly there is $r_4 = (r_1[X], r_2[Y], r_1[Z]) \in R$. Then $X \overset{\lambda}{\twoheadrightarrow} Y$ holds in R. The converse is also easily shown.

The reader can verify that the inference rules, as has been done for FD and MVD [2] satisfy for λFD and λMVD, too. The inference rules are presented in following:
X,Y,Z and V are subsets of W.

<u>λFD inference rules</u>

λFD1 : if $Y \subseteq X$ then $X \overset{\lambda}{\rightarrow} Y$.

λFD2 : if $Z \subseteq V$ and $X \overset{\lambda}{\rightarrow} Y$ then $XV \overset{\lambda}{\rightarrow} YZ$.

λFD3 : if $X \overset{\lambda}{\rightarrow} Y$ and $Y \overset{\lambda}{\rightarrow} Z$ then $X \overset{\lambda}{\rightarrow} Z$.

λFD4 : if $X \overset{\lambda}{\rightarrow} Y$ and $YV \overset{\lambda}{\rightarrow} Z$ then $XV \overset{\lambda}{\rightarrow} Z$.

λFD5 : if $X \overset{\lambda}{\rightarrow} Y$ and $X \overset{\lambda}{\rightarrow} Z$ then $X \overset{\lambda}{\rightarrow} YZ$.

λFD6 : if $X \overset{\lambda}{\rightarrow} YZ$ and $X \overset{\lambda}{\rightarrow} Y$ then $X \overset{\lambda}{\rightarrow} Z$.

<u>λMVD inverence rules</u>

λMVD0 : $X \overset{\lambda}{\twoheadrightarrow} Y$ iff $X \overset{\lambda}{\twoheadrightarrow} W \setminus Y$.

λMVD1 : if $Y \subseteq X$ then $X \overset{\lambda}{\twoheadrightarrow} Y$.

λMVD2 : if $Z \subseteq V$ and $X \overset{\lambda}{\twoheadrightarrow} Y$ then $XV \overset{\lambda}{\twoheadrightarrow} YZ$.

λMVD3 : if $X \overset{\lambda}{\twoheadrightarrow} Y$ and $Y \overset{\lambda}{\twoheadrightarrow} Z$ then $X \overset{\lambda}{\rightarrow} Z \setminus Y$.

λMVD4 : if $X \overset{\lambda}{\twoheadrightarrow} Y$ and $XV \overset{\lambda}{\twoheadrightarrow} Z$ then $XV \overset{\lambda}{\twoheadrightarrow} Z \setminus YV$.

λMVD5 : if $X \overset{\lambda}{\twoheadrightarrow} Y$ and $X \overset{\lambda}{\twoheadrightarrow} Z$ then $X \overset{\lambda}{\twoheadrightarrow} YZ$.

λMVD6 : if $X \overset{\lambda}{\twoheadrightarrow} Y$ and $X \overset{\lambda}{\twoheadrightarrow} Z$ then $X \overset{\lambda}{\twoheadrightarrow} Y \cap Z$, $X \overset{\lambda}{\twoheadrightarrow} Y \setminus Z$, $X \overset{\lambda}{\twoheadrightarrow} Z \setminus Y$.

<u>λFD-λMVD inference rules</u>

λFD-λMVD1 : if $X \overset{\lambda}{\rightarrow} Y$ then $X \overset{\lambda}{\twoheadrightarrow} Y$.

λFD-λMVD2 : if $X \overset{\lambda}{\twoheadrightarrow} Z$ and $Y \overset{\lambda}{\rightarrow} Z'$, $(Z' \subseteq Z,\ Y \cap Z = \emptyset)$

then $X \overset{\lambda}{\rightarrow} Z'$.

λFD-λMVD3 : if $X \overset{\lambda}{\twoheadrightarrow} Y$ and $XY \overset{\lambda}{\twoheadrightarrow} Z$ then $X \overset{\lambda}{\rightarrow} Z \setminus Y$.

Let $\mathcal{B}$ be a family of all n-ary decompositions of the x-relation R over W (denoted by $(X_1,\dots,X_n)$). This family has the following properties.

Theorem 5.1. [6].

Let R an x-relation over W. $X_i \subseteq W$, $i = \overline{1,n}$, $\bigcup_{i=1}^{n} X_i = W$. The family $\mathcal{B}$ of all n-ary decompositions of the x-relation R satisfies the following conditions:

1. $(\emptyset,\dots,\emptyset,W) \in \mathcal{B}$.
2. If $(X_1,\dots,X_n) \in \mathcal{B}$ then $(X_{\pi(1)},\dots,X_{\pi(n)}) \in \mathcal{B}$ where $\pi: \{1,\dots,n\} \to \{1,\dots,n\}$ is a permutation.
3. If $(X_1,\dots,X_n) \in \mathcal{B}$ and $X_i \subseteq Y \subseteq W$, $i = \overline{1,n}$, then $(X_1,\dots,X_{i-1},Y,X_{i+1},\dots,X_n) \in \mathcal{B}$.
4. If $(X_1,\dots,X_n) \in \mathcal{B}$ and $X_i \subseteq X_j$, $i \neq j$ then $(X_1,\dots,X_{i-1},\emptyset,X_{i+1},\dots,X_j,\dots,X_n) \in \mathcal{B}$.
5. If $(X_1,\dots,X_n)$ and $(Y_1,\dots,Y_n) \in \mathcal{B}$ with $Y_1 \cap Y_i = X_i$, $i = \overline{2,n}$ and $Y_i \cap Y_j \subseteq X_i \cap X_j$, $i \neq j$, $i,j = \overline{2,n}$ then $(X_1 \cap Y_1, Y_2,\dots,Y_n) \in \mathcal{B}$.

Proof.

Like the proof in [6].

5.2. Ternary degenerate decomposition of an x-relation

An interesting class of decompositions of an x-relation which plays an important role in the design process of a database, is the acyclic decomposition [3].

In [5] the author has investigated the general properties of a full family of all ternary decompositions of a relation. In this subsection we will only consider the class of ternary acyclic decompositions of an x-relation. (The following results are correct for a ternary acyclic decomposition of an usual relation). Let (X,Y,Z) be a ternary decomposition of an x-relation R(W). If X,Y,Z are no-empty subsets of W with $XYZ = W$, $X \neq Y \neq Z$ and $X \cap Y \neq \emptyset$, $Y \cap Z \neq \emptyset$ and $X \cap Z = \emptyset$ then this decomposition is said acyclic.

Proposition 5.2. [3,6]

Let R be an x-relation over W. X,Y,Z are no empty subsets of W with $XYZ = W$ and $X \cap Y \neq \emptyset$, $Y \cap Z \neq \emptyset$, $X \cap Z = \emptyset$. (X,Y,Z) is a ternary acyclic decomposition of x-relation R iff the following data dependencies are satisfied in R:

$$X \cap Y \overset{\lambda}{\twoheadrightarrow} X,\ (X \cap Y)(Y \cap Z) \overset{\lambda}{\twoheadrightarrow} Y \text{ and } (Y \cap Z) \overset{\lambda}{\twoheadrightarrow} Z.$$

Proof.

It is easy to verify.

It follows directly from definition 5.2 and theorem 5.1 the following:

Proposition 5.3.

If (X,Y,Z) is a ternary acyclic decomposition of an x-relation R over W, then (XY,Z),(X,YZ),(XY,YZ) are binary decompositions of R and (X,Y),(Y,Z) binary decompositions of projections R[XY] and R[YZ], respectively.

To capture more the semantics of data,we use here the concept of degenerate multivalued dependencies [12] for determining which join of relations can be updated by insertion or deletion of a tuple without other tuples entering or leaving the join.

Definition 5.5. [12]

A λ-multivalued degenerate dependency $X \overset{\lambda}{\twoheadrightarrow} Y|Z$ holds in an x-relation $R(W)$ with $Z = W \setminus XY$ if for every pair of tuples $r_1, r_2 \in R$, $r_1[X] \overset{\lambda}{=} r_2[X]$, either $r_1[Y] \overset{\lambda}{=} r_2[Y]$ or $r_1[Z] \overset{\lambda}{=} r_2[Z]$.

The definition 5.5 can be reformulated in the other form of the corresponding binary decomposition as follows.

Definition 5.6.

Let R be an x-relation over W and (X,Y) be a binary decomposition of R.
(X,Y) is binary degenerate decomposition of R if there exist two x-relations R_1 and R_2 such that $R = R_1 \cup R_2$, $X \cap Y \overset{\lambda}{\to} X$ holds in R_1 and $X \cap Y \overset{\lambda}{\to} Y$ holds in R_2 with $R_1[X \cap Y] \cap R_2[X \cap Y] = \emptyset$.

In [1] the authors have studied the effect of type of binary degenerate multivalued dependency for the update problem. For our purpose, the most important of this result is that a relation R can be deletion- or insertion- viable if and only if a certain degenerate decomposition holds in R.

Attempts to generalize this result to decompositions of an x-relation (in meaning of λ-equivalence) into more than two components we introduce a new concept called ternary degenerate decomposition.

Definition 5.7.

A ternary acyclic decomposition (X,Y,Z) of an x-relation R over W is degenerate if there exist two x-relations R_1 and R_2 such that $R = R_1 \cup R_2$, $X \cap Y \overset{\lambda}{\to} YZ$ holds in R_1 and

$Y \cap Z \overset{\lambda}{\to} XY$ holds in R_2 with $R_1[X \cap Y] \cap R_2[X \cap Y] = \emptyset$ or $R_1[Y \cap Z] \cap R_2[Y \cap Z] = \emptyset$.

In the following it is assumed that (X,Y,Z) is a ternary acyclic decomposition of $R(W)$. This decomposition has some properties as follows.

Proposition 5.4.

If (X,Y,Z) is a ternary degenerate decomposition of R, then (X,YZ), (XY,Z) and (XY,YZ) are binary degenerate decompositions of R; (X,Y) and (Y,Z) are binary degenerate decompositions of the corresponding projection $R[XY]$ and $R[YZ]$.

Proof.

It must be show that (X,YZ) is a binary degenerate decomposition. Since (X,Y,Z) is degenerate, then there exist R_1 and R_2 such that $R = R_1 \cup R_2$ and $X \cap Y \overset{\lambda}{\to} YZ$ in R_1, $Y \cap Z \overset{\lambda}{\to} XY$ in R_2. W.l.o.g. it is assumed that $R_1[X \cap Y] \cap R_2[X \cap Y] = \emptyset$.

First we have to show that (X,YZ) is a binary degenerate decomposition of R.

We have $X \cap Y \overset{\lambda}{\to} YZ$ that holds in R_1. It remains only to show that $X \cap Y \overset{\lambda}{\to} X$ holds in R_2.

Since (X,YZ) is a binary decomposition and $R_1[X \cap Y] \cap R_2[X \cap Y] = \emptyset$, the following data dependencies hold in R_2:

$$X \cap Y \overset{\lambda}{\twoheadrightarrow} X, \quad Y \cap Z \overset{\lambda}{\to} XY.$$

Hence $Y \cap Z \overset{\lambda}{\to} X$.

After the applications of inference rule λFD- MVD2 we have $X \cap Y \overset{\lambda}{\to} X$.

In order to show that (X,YZ) is degenerate under the above assumptions, we must construct R_1' and R_2' as follows. Since (XY,Z) is a binary decomposition then it satisfies the following conditions: for any pair of tuples $r_1, r_2 \in R$ with

$r_1[Y \cap Z] \stackrel{\lambda}{=} r_2[Y \cap Z]$, there exists a tuple $r \in R$ such that $r[XY] \stackrel{\lambda}{=} r_1[XY]$ and $r[Z] \stackrel{\lambda}{=} r_2[Z]$.

Three cases can happen.

- Case 1: if $r_1, r_2 \in R_1$ then it is easy to see that $r \in R_1$, too.
- Case 2: of $r_1, r_2 \in R_2$ then it is easy to see that $r \in R_2$, too.
- Case 3: w.l.o.g. it is assumed that $r_1 \in R_1$ and $r_2 \in R_2$.

If $r \in R_2$ then $r[XY] \stackrel{\lambda}{=} r_1[XY]$ and $r[Z] \stackrel{\lambda}{=} r_2[Z]$.
It follows that $r[X \cap Y] \stackrel{\lambda}{=} r_1[X \cap Y]$, which contradicts to the fact that $R_1[X \cap Y] \cap R_2[X \cap Y] = \emptyset$. It remains only the case that $r \in R_1$.
Since $r[XY] \stackrel{\lambda}{=} r_1[XY]$ and $X \cap Y \stackrel{\lambda}{\to} YZ$ holds in R_1, then $r \equiv r_1$ in R_1. Hence $r_1[Z] \stackrel{\lambda}{=} r_2[Z]$.
This shows that if there exists a tuple $r_1 \in R_1$ such that $r_1[Y \cap Z] \stackrel{\lambda}{=} r_2[Y \cap Z]$ where $r_2 \in R_2$, then it must satisfy $r_1[Z] \stackrel{\lambda}{=} r_2[Z]$. All such tuples of R_2 can be grouped to a set T. Then two x-relations R_1' and R_2' can be constructed as follows:

$$R_1' = R_1 \cup T, \quad R_2' = R_2 \backslash T.$$

We have also $R_1'[Y \cap Z] \cap R_2'[Y \cap Z] = \emptyset$ and the data dependencies $Y \cap Z \stackrel{\lambda}{\to} Z$ and $Y \cap Z \stackrel{\lambda}{\to} XY$ holds respectively in R_1' and R_2'. Hence (XY,Z) is a degenerate decomposition of R.

All remaining decompositions are easy to verify to be degenerate of R.

Proposition 5.5.

If (XY,Z) and (X,YZ) are degenerate decompositions of

R, then (X,Y) and (Y,Z) are degenerate decompositions of $R[XY]$ and $R[YZ]$, respectively.

Proof. Trivial.

Propositions 5.6.

If (XY,Z) is a degenerate decomposition of R and (X,Y) is a degenerate decomposition of $R[XY]$, then (X,YZ) is a degenerate decomposition of R and (Y,Z) is a degenerate decomposition of $R[YZ]$.

Proof.

Since (XY,Z) is a degenerate, then there exists two x-relations R_1 and R_2 with $R = R_1 \cup R_2$ such that $Y \cap Z \overset{\lambda}{\rightarrow} XY$ holds in R_1 and $Y \cap Z \overset{\lambda}{\rightarrow} Z$ holds in R_2 and $R_1[Y \cap Z] \cap R_2[Y \cap Z] = \emptyset$.

We want to construct now two x-relations R_1' and R_2' which satisfy all conditions of (X,YZ) to be degenerate. Since (X,YZ) is a binary decomposition, then for any pair of tuples $r_1, r_2 \in R$ with $r_1[X \cap Y] \overset{\lambda}{=} r_2[X \cap Y]$, there exists $r \in R$ such that $r[X] \overset{\lambda}{=} r_1[X]$ and $r[YZ] \overset{\lambda}{=} r_2[YZ]$.

There are three cases:

Case 1.

If $r_1, r_2 \in R_1$ then $r \in R_1$. Since, if $r \in R_2$ then $r[YZ] \overset{\lambda}{=} r_2[YZ]$. Therefore $r[Y \cap Z] \overset{\lambda}{=} r_2[Y \cap Z]$. It is contrary to $R_1[Y \cap Z] \cap R_2[Y \cap Z] = \emptyset$.

Case 2.

Similarly, it is shown that if $r_1, r_2 \in R_2$ then $r \in R_2$.

Case 3.

W.l.o.g. it is assumed that $r_1 \in R_1$ and $r_2 \in R_2$.

If $r \in R_1$ then $r[Y \cap Z] \stackrel{\lambda}{=} r_2[Y \cap Z]$. It is contrary to $R_1[Y \cap Z] \cap R_2[Y \cap Z] = \emptyset$.

There is also only the case that $r_1 \in R_1$ and $r, r_2 \in R_2$. Because (X,Y) is degenerate decomposition in $R[XY]$, the tuples r_1, r_2 satisfy $r_1[X \cap Y] \stackrel{\lambda}{=} r_2[X \cap Y]$, then $r_1[X] \stackrel{\lambda}{=} r_2[X]$ or $r_1[Y] \stackrel{\lambda}{=} r_2[Y]$. But from degenerate decomposition (XY, YZ) we get $r_1[Y] \neq r_2[Y]$. It follows that $r_1[X] \stackrel{\lambda}{=} r_2[X]$. This means, that $r \equiv r_2 \in R_2$. The last case show that if there exists a tuple $r \in R_2$ such that $r[X \cap Y] \stackrel{\lambda}{=} r'[X \cap Y]$, $r' \in R_1$, it must satisfy $r[X] \stackrel{\lambda}{=} r'[X]$. All such tuples of R_2 can be grouped to a set T. Then two x-relations can be constructed as follows:

$$R_1' = R_1 \cup T \quad \text{and} \quad R_2' = R_2 \setminus T.$$

We have also $R = R_1' \cup R_2'$, $R_1'[X \cap Y] \cap R_2'[X \cap Y] = \emptyset$. From the construction of R_1' and R_2' as above, it is easily seen that the data dependency $X \cap Y \stackrel{\lambda}{\rightarrow} X$ holds in R_1' and the data dependency $X \cap Y \stackrel{\lambda}{\rightarrow} YZ$ holds in R_2'. Therefore (X, YZ) is a degenerate decomposition of R.

REFERENCES

[1] W.W. Armstrong, C. Delobel: Decompositions and functional dependencies in relations, ACM TODS 5,4 (1980), 404-430.

[2] C. Beeri, R. Fagin, J.H. Howard: A complete axiomatization for functional and multivalued dependencies in database relations, Proc. 1977 ACM SIGMOD, Toronto, 47-61.

[3] C. Beeri, R. Fagin, D. Maier, M. Yannakakis: On the desirability of acyclic database schemes, J.ACM 30,3(1983), 479-513.

[4] E.F. Cood: A relational model for large schared data banks, C.ACM 13,6(1970) 377-387.

[5] Le Tien Vuong: Untersuchung zur ternaeren Dekomposition einer Relation und zur Anwendung der unscharfen Mengen im CRM, Diss. TU-Dresden 1983.

[6] Le Tien Vuong: Über n-fache Dekomposition einer Relation im Coddschen Relationenmodell, MTA SZTAKI Közlemenyek, 31/1984, 95-113.

[7] Le Tien Vuong and Ho Thuan: Retrieval from fuzzy database by fuzzy relational algebra, to appear in MTA SZTAKI Közlemények 37/1987.

[8] Y.E. Lien: Multivalued dependencies with null values in relational databases, 5th Int. Conf. VLDB, Rio de Janeiro, 1979, 155-168.

[9] C.V. Negoita, D.A. Ralesa: Applications of fuzzy sets to systems analysis, Birk. Verlag, 1975.

[1o] H. Prade, C. Testemale: Traitement de questions vagues dans une base de données imprécises 1,2, L'Infor. Process. 27(1984) 65-84, 28(1984) 49-66.

[11] R. Fagin: Multivalued dependencies and a new Normal Form for relational databases, ACM TODS 2,3(1977), 262-278.

[12] Y. Sagiv, C. Delobel, D.S. Jr. Parker, R. Fagin: An equivalence between relational database dependencies and a fragment of propositional logic, J.ACM 28,3(1981), 435-453.

[13] V. Tahani: A conceptual framework for fuzzy query processing-a step toward very intelligent database systems, Inf. Proc. and Manag., 13(1977),289-303.

[14] M. Umano, M. Mizimoto, K. Tanaka: Fuzzy database systems, Proc. of Working Conf. on database Engineering (13th IBM Computer Science Symp. IBM Amagi Homestest, Japan 1979, 17-19.

[15] C. Zaniolo: Database relations with null values, J. of Computer and Syst. Sciences 28(1984) 142-166.

[16] L.A. Zadeh: The concept of a linguistic variable and its applications to approximate reasoning, Inf. Science 8(1975/ 199-248 and 301-357.

Обобщение реляционных баз данных применяя теорию "фаззи" множеств и лингвистических переменных

Лие Тиен Вуонг, Хо Тхуан

Резюме

В реальном мире существуют данные,содержание которых очень неточное /например, "он молодой"/. Для разработки таких данных и для построения теоретических основ таких баз данных уже применяется теория "фаззи" множеств. Другой подход - ввести в базу данных переменные нового типа, т.н. лингвистические переменные. В статье разработана теория баз данных, используя оба метода.

A "FUZZY" HALMAZ-ELMÉLET ILL. "LINGVISZTIKAI" VÁLTOZÓK ÁLTAL KIBŐVITETT RELÁCIÓS ADATBÁZISOK

Lien Tien Vuong, Ho Thuan

Összefoglaló

A valós világban olyan adatok is vannak, amelyek tartalma eléggé pontatlan (pl. "fiatal"). Az ilyen adatok feldolgozására "fuzzy" ("elmosódott") halmazelméleten alapuló ujfajta adatbázis-elméletet dolgoztak ki. Egy másik módszer az u.n. "lingvisztikai" ("nyelvészeti") változók bevezetése. A cikkben a két módszer alapján egy újfajta relációs adatbázis-elmélet részletes ismertetése van.