AN APPROACH TO EXTENDING THE RELATIONAL DATABASE MODEL FOR HANDLING INCOMPLETE INFORMATION AND DATA DEPENDENCIES

HO THUAN, HO CAM HA

Abstract. In this paper we propose a new approach to extending the relational database model. This approach is based on the concept of similarity based fuzzy relational database and somewhat of new viewpoint on redundancy. It is shown that, in such an extended database model, we can capture imprecise, uncertain information. The formal definition of fuzzy functional and multivalued dependencies in this study allows a sound and complete set of inference rules. This paper describes an ongoing work. We state some open problems to be solved in order to render our approach more operational.

Tóm tắt. Bài báo đề xuất một cách tiếp cận mới để mở rộng mô hình cơ sở dữ liệu quan hệ. Cách tiếp cận này dựa trên khái niệm cơ sở dữ liệu mờ tương tự và một quan điểm mới về dư thừa dữ liệu. Với mô hình cơ sở dữ liệu như vậy có thể nắm bắt được những thông tin không chính xác, không chắc chắn. Định nghĩa về phụ thuộc hàm mờ và phụ thuộc đa trị mờ trong bài báo cho một tập các luật suy dẫn xác đáng và đầy đủ.

1. INTRODUCTION

Database systems have been extensively studied since Codd [3] proposed the relational data model. Such database systems do not accept uncertain and imprecise data. In fact, the value of an object's attribute may be completely unknown, incompletely known (i.e., only a subset of possible values of the attribute is known), or uncertain (e.g. a probability or possibility distribution for its value is known). In addition, the attribute may not be applicable to some of the objects being considered and, in certain cases, we may not known whether the value even exists, or not. Many approaches to that problem have been proposed. One of them is "A fuzzy representation of data for relational database" [2], which is suggested by P. Buckles and E. Petry. In [2] a structure for representing inexact information in the form of a relational database is presented. The structure differs from ordinary relational database in two important respects: value of an attribute of an object need not be single value and a similarity relation is required for each domain set of the database. In a fuzzy database proposed by these authors, a tuple is redundant if it can be merged with another through the set union of corresponding domain values. The merging of tuple, however, is subject to constraints on some similar thresholds. Within this conception, in a fuzzy relation with no redundant tuples and each domain similarity relation formulated according to T1 transitivity, each tuple represents information of an object, and each value of an attribute (called domain value) consists of one or more elements from the domain base set. At this point, there is an emphatic notice that elements of each domain value must be similar enough to each other (i.e. similarity degree of every couple of elements is not less than the given threshold).

The work reported here is quite distinct from that of P. Buckles and E. Petry in that the elements of each domain value are not required to be similar enough according to the threshold. This idea allows each domain value to contain elements, which even are not very similar and represent the possibilities that can be happened. Therefore, to model a relational database by using this approach will preserve not only the exact information but also the nuances of fuzzy uncertainty.

This paper is organized as follows. Notations and basic definitions related to fuzzy relational data model and similarity relation, are reviewed in Section 2 to get an identical understanding of terminology. A new definition about tuple redundant is presented in Section 3. Section 4 contains

definition of functional dependency in this scene. The soundness and completeness of the set of axioms, which is similar with Amstrong's axioms in the traditional relational database, will be proved in this section. In Section 5, we propose a formal definition of fuzzy multivalued dependency and the inference rules.

2. BACKGROUND

First, similarity relations are described as defined by Zadeh [9]. Then the basic concepts of fuzzy relational database model are reviewed.

Similarity relations are useful for describing how similar two elements from the same domain are.

Definition 2.1. ([5]) A similarity relation, $s_D(x, y)$, for a given domain D, is a mapping of every pair of elements in the domain onto the unit interval [0, 1] with the three following properties, $\forall x, y, z \in D$:

1.	Reflexivity	$S_D(x,x)=1$
2.	Symmetry	$S_D(x,y) = S_D(y,x)$
3.	Transitivity	$S_D(x,z) \ge \underset{y}{\operatorname{Max}}(\operatorname{Min}[S_D(x,y), S_D(y,z)])$ (T1)

or 3'. Transitivity

$$S_D(x,z) = \operatorname{Max}_y([S_D(x,y) * S_D(y,z)])$$
 (T2)

where * is arithmetic multiplication)

For each domain j in a relational database, a domain base set D_j is understood. Domains for fuzzy relational databases will be either discrete scalars or discrete numbers drawn from either a finite or infinite set. A domain value d_{ij} , where i is the tuple index, is defined to be a subset (not empty) of its domain base set D_j . Let 2^{D_j} denote a set of any non-null member of the powerset of D_j .

Definition 2.2. ([2]) A fuzzy relation, r, is a subset of the set cross product $2^{D_1} \times \cdots \times 2^{D_m}$.

Definition 2.3. ([2]) A fuzzy relation tuple, t, is any member of $2^{D_1} \times \cdots \times 2^{D_m}$.

An arbitrary tuple is of the form $t_i \in r$, $t_i = (d_{i1}, d_{i2}, \dots, d_{im})$, $d_{ij} \subseteq D_j$. For example:

Name	Car_color	Job	
$\{\mathrm{John}\}$	{green, blue, pink}	{doctor, physician, dentist, farmer}	

3. REDUNDANCY AND DETERMINANCY PROPERTIES

In a nonfuzzy database, a tuple is redundant if it is exactly the same as another tuple. In fuzzy database of P.Buckles and E.Petry [2], a tuple is redundant if it can be merged with another without violating

LEVEL
$$(D_j) = \text{THRES}(D_j), j = 1, 2, ..., m$$
, where $\text{THRES}(D_j) = \min_i \{\min_{x,y \in d_{ij}} [s(x,y)]\}$ [2]

In a given domain D_j , $x, y \in D_j$, if $s(x, y) \ge \text{LEVEL}(D_j)$ then we write down $x \sim y$. Obviously, \sim is a binary relation on D_j .

Lemma 3.1. ~ is an equivalence relation.

Proof. $\forall x \in D_j$, s(x, x) = 1, so $s(x, x) \ge \text{LEVEL}(D_j)$, we have $x \sim x$.

Symmetry property of \sim relation is easily implied from the symmetry property of a similarity measure. $\forall x, y, z \in D_j$, if $s(x, y) \ge \text{LEVEL}(D_j)$ and $s(y, z) \ge \text{LEVEL}(D_j)$, from (T1) transitivity we have $s(x, z) \ge \text{LEVEL}(D_j)$.

Thus, \sim is an equivalence relation and induces a unique partition in D_j .

In a fuzzy relational scheme suggested by Buckles and Petry [2], each domain value may consist of many elements, all of which belong to the same equivalence class partitioned by the \sim relation.

According to these authors, two tuples are redundant to each other if on every attribute, the domain value of each tuple includes representatives of the same equivalence class. To a certain meaning, if we consider an equivalence class (of the ~ relation) as a branch of possibilities that may happen, the model of P. Buckles and E. Petry will allow only to capture information of the objects, of which the known information about each attribute belongs to only one branch of possibilities. The branch of possibilities mentioned here is considered to be shown by values, which are, although not equal to each other, but closed enough to each other according to the measure of a similarity relation. However, in fact there can be uncertain information about an object, on an attribute of that there are many possibilities which are far different to each other. In the above example, John may be a doctor, a physician, a dentist (or any position in medical profession), but John may be also a farmer. John has a green car, or a pink one, but he may have two cars, one is blue and the other is pink. And it is not excluded that John has all the three cars which are green, blue and pink. If a group of possibility branches is considered necessary to keep as it identifies a full information in this case, the model in [2] should be expanded, and we have tried to do this. Suppose that with each Dj there is a LEVEL(Dj) for an identified similarity on this domain, two tuples are said to be redundant to each other if they have the same group of possibilities on each attribute.

Definition 3.1. In fuzzy relation r, two tuples $t_i = (d_{i1}, d_{i2}, \dots, d_{im})$ and $t_k = (d_{k1}, d_{k2}, \dots, d_{km})$, $i \neq k$ are redundant if

 $\forall x \in d_{ij} \ \exists x' \in d_{kj} : x \sim x', \ \forall j = 1, 2, ..., m \ \text{and vice versa, i.e.}$

 $\forall x \in d_{kj} \ \exists x' \in d_{ij} : x \sim x', \ \forall j = 1, 2, \dots, m.$

As t_i and t_k are equitable in the above definition, the notation $t_i \approx t_k$ is used to denote that t_i and t_k are redundant.

Lemma 3.2. \approx is an equivalence relation on the fuzzy relation r.

Proof. It is clear that, for every tuple t_i of r, $t_i \approx t_i$ from reflexivity of \sim relation.

Obviously, if $t_i \approx t_k$ then $t_k \approx t_i$.

Suppose that $t_i \approx t_k$ and $t_k \approx t_h$. Consider arbitrary domain D_j , if $x \in d_{ij}$ then $\exists x' \in d_{kj} : x \sim x'$ (from $t_i \approx t_k$). Since $x' \in d_{kj}$, we have $\exists x'' \in d_{hj} : x' \sim x''$ (from $t_k \approx t_h$). We also have $x \sim x''$ by transitivity of \sim relation. Similarly, if $x \in d_{hj}$ we have $\exists x'' \in d_{ij} : x \sim x''$.

Thus, redundant (\approx) is an equivalence relation on R and induces a unique partition in r.

An example of a fuzzy relation with similarity relations:

r1	Name	Car_color	Job
t_1	John	green, blue, pink	actor, teacher
t_2	Johan	black, magent	aconductor, instructor
t_3	Elina	white, pink	artist
t_4	Melia	pink, light_milk	artist
t_5	Tom	black, red	pilot

Fig. 1. A fuzzy relation

If it is assumed that LEV(Name) = 0.6 then \sim relation partitions Dom (Name) by three equivalence classes:

{John, Johan}; {Elina, Melina}; {Tom}

It is also assumed that LEV(Car_color) and LEV(Job) are given such that

Dom(Car_color) and Dom(Job) are partitioned as follow

{{green, blue, black}, {pink, magenta, red}, {white, light_milk}}

{{actor, conductor, artist}, {teacher, instructor}, {pilot}}

Thus in r1 above, t_1 is redundant for t_2 and t_3 is redundant for t_4 .

	John	Johan	Elina	Melina	Tom
John	1.0	0.6	0.0	0.0	0.0
Johan	0.6	1.0	0.0	0.0	0.0
Elina	0.0	0.0	1.0	0.8	0.0
Melina	0.0	0.0	0.8	1.0	0.0
Tom	0.0	0.0	0.0	0.0	1.0

Fig. 2. Similarity relation for Dom(Name)

4. FUZZY FUNCTIONAL DEPENDENCY AND A SET OF SOUND AND COMPLETE INFERENCE RULES

Let r is a fuzzy relation with m attributes, these according to m domains D_1, D_2, \ldots, D_m , we said that r is an instance of R, which is called a relation scheme on $U, U = \{A_1, A_2, \ldots, A_m\}$. Suppose that X is a set of attributes $(X \subseteq U)$, two tuples $t_1, t_2 \in r$, $t_1 = (d_{11}, d_{12}, \ldots, d_{1m})$ and $t_2 = (d_{21}, d_{22}, \ldots, d_{2m})$, we said t_1, t_2 are redundant each other on X and write $t_1[X] \approx t_2[X]$ if $\forall x \in d_{1j} \ \exists x' \in d_{2j} : x \sim x'$, and vice versa, i.e. $\forall x \in d_{2j} \ \exists x' \in d_{1j} : x \sim x', \ \forall j : A_i \in X$.

Definition 4.1. A fuzzy functional dependency $X \approx Y$ is said to be hold in a fuzzy relation r if for every pairs of tuple $t_1, t_2 \in r$:

$$t_1[X] \approx t_2[X]$$
 implies that $t_1[Y] \approx t_2[Y]$.

In what follows we assume that we are given a fuzzy relational schema with set of attribute U, the universal set of attributes, and a set of fuzzy functional dependencies F involving only attributes in U. The inference rules, which similar with Amstrong's axioms are:

FFD1: Reflexivity

If $Y \subseteq X$ then $X \approx \rightarrow Y$

FFD2: Augmentation

If $X \approx Y$ holds, then $XZ \approx YZ$ holds

FFD3: Transitivity

If $X \approx Y$ and $Y \approx Z$ hold, then $X \approx Z$ holds

Lemma 4.1. The set of FFD axioms (FFD1-FFD3) are sound. That is, if $X \approx Y$ is deduced from F using the axioms, then $X \approx Y$ is true in any relation in which the dependencies of F are true.

Proof.

(FFD1) The reflexivity axiom is clear sound.

(FFD2) Suppose $t_1, t_2 \in r$ such that $t_1[XZ] \approx t_2[XZ]$ (1) then by definition of " \approx " we have $t_1[X] \approx t_2[X]$. From $X \approx Y$ we have $t_1[Y] \approx t_2[Y]$ (2)

- (1) means $\forall x \in d_{1j} \ \exists x' \in d_{2j} : x \sim x'$, and vice versa $\forall j : D_j \in XZ$.
- (2) means $\forall x \in d_{1j} \ \exists x' \in d_{2j} : x \sim x'$, and vice versa $\forall j : D_j \in Y$.

So we have

$$\forall x \in d_{1j} \ \exists x' \in d_{2j} : x \sim x'$$
, and vice versa $\forall j : D_j \in YZ$.

It means $XZ \approx YZ$.

(FFD3) If
$$t_1[X] \approx t_2[X]$$
 then we have $t_1[Y] \approx t_2[Y]$ from $X \approx Y$ and $t_1[Z] \approx t_2[Z]$ from $Y \approx Z$.

The following inference axioms are infered from the above axioms

FFD4: Union If $X \approx Y$ and $X \approx Z$ hold, then $X \approx YZ$ holds. FFD5: Decomposition If $X \approx YZ$ holds, then $X \approx YZ$ hold.

FFD6: Pseudotransitivity If $X \approx Y$ and $YW \approx Z$ hold, then $XW \approx Z$ holds.

Procedure of proof for the completeness of above inference axioms is very similar to the classical case.

Theorem 4.1. The set of axioms (FFD1-FFD2) are sound and complete.

5. FUZZY MULTIVALUED DEPENDENCY AND SET OF INFERENCE RULES

In the fuzzy paradigm, let R be a relation scheme and let X and Y be subsets of R. In a relation r, an instance of R, for X-value x we define

 $X_r(x) = \{x' | \exists t \in r, \text{ such that } t[X] = x', x \approx x'\}.$

r2

 $Y_r(x) = \{y | \exists t \in r, \text{ such that } t[X] \in X_r(x), t[Y] = y\}.$

Let Z = R - XY. It is clear that $Y_r(x)$ is independent of Z-values. We say that $Y_r(x)$ is equivalent to $Y_r(xz)$ if for every y of one, there is existing y' of the other such that $y \approx y'$ and vice versa. The fuzzy equivalence of two set Y-value $(Y_r(x))$ and $Y_r(xz)$ can be repersented as $Y_r(x) \cong Y_r(xz)$.

Definition 5.1. A fuzzy multivalued dependency (FMVD) m on a scheme R, is a statement $m: X \approx \longrightarrow Y$, where X, Y are subsets of R. Let Z = R - XY. A relation r on the scheme R obeys the FMVD $m: X \approx \longrightarrow Y$ if for every XZ-value xz that appears in r we have $Y_r(x) \cong Y_r(xz)$.

Example:

X (Degree)	Y (Courses)	Z (Student)
a, b, c	g, h	z 1
a', c'	g', i	z 2
a, c'	g, i'	z1'
a', c	g', h'	z2'

Fig. 3. A fuzzy relation

$$egin{aligned} x1 &= \{a,b,c\}, & X_r(x1) &= \{\{a,b,c\}, \{a',c'\}, \{a,c'\}, \{a',c\}\}\} \ & Y_r(x1) &= \{\{g,h\}, \{g',i\}, \{g,i'\}, \{g',h'\}\} \ & Y_r(x1z1) &= \{\{g,h\}, \{g,i'\}\} \end{aligned}$$

It is assumed that:

$$a \sim b \sim a'$$
 ; $c \sim c'$;
 $g \sim g'$; $h \sim h'$; $i \sim i'$;
 $z1 \sim z1'$; $z2 \sim z2'$.

Therefore $\{g', i\} \cong \{g, i'\},$ $\{g', h'\} \cong \{g, h\},$

so $Y_r(x1) \cong Y_r(x1z1)$, and by similar reasoning we must have $Y_r(x1) \cong Y_r(x1z2)$.

We say fuzzy multivalued $X \approx \longrightarrow Y$ is satisfied in r2.

We now propose the set of fuzzy functional and multivalued dependencies inference rules over a set of atributes U. The first three for fuzzy functional dependencies are repeat here.

A1: Reflexivity for fuzzy functional dependencies (FFD)

If $Y \subseteq X$ then $X \approx Y$.

A2: Augmentation for FFD

If $X \approx Y$ holds, then $XZ \approx YZ$ holds.

A3: Transitivity for FFD

If $X \approx Y$ and $Y \approx Z$ hold, then $X \approx Z$ holds.

A4: Complementation for fuzzy multivalued dependencies (FMVD)

If $X \approx \longrightarrow Y$ holds, then $X \approx \longrightarrow Z$, where Z = R - XY.

A5: Augmentation for FMVD

If $X \approx \longrightarrow Y$ holds, then $XZ \approx \longrightarrow YZ$ holds.

A6: Transitivity for FMVD

If $X \approx \longrightarrow Y$ and $Y \approx \longrightarrow Z$ hold then $X \approx \longrightarrow (Z - Y)$ holds.

Last two axioms that relate fuzzy functional and fuzzy multivalued dependencies are also similar to classical cases.

A7: If $X \approx Y$ holds, then $X \approx Y$.

A8: If $X \approx \longrightarrow Y$ holds, $Z \subseteq Y$, $W \cap Y = \emptyset$, and $W \approx \longrightarrow Z$, then $X \approx \longrightarrow Z$ holds.

Lemma 5.1. The set of axioms (A1-A8) are sound. That is, if the fuzzy dependency (FFD or FMVD) is deduced from a set of FFDs and FMVDs, G, using the axioms, then it is true in any relation in which the dependencies of G are true.

Proof. By Lemma 4.1, the axioms A1-A3 is sound.

(A4) Complementation for fuzzy multivalued dependencies (FMVD)

If $X \approx \longrightarrow Y$ holds, then $X \approx \longrightarrow Z$, where Z = R - XY.

We shall prove that, if for every XZ-value xz that appears in r we have $Y(x) \cong Y(xz)$ then $Z(x) \cong Z(xy)$ for every XY-value xy that appears in r. Obviously, $Z(xy) \subseteq Z(x)$. Therefore, we only need to show

$$\forall z_0(Z(x) \exists z' \in Z(xy) : z_0 \approx z'. \tag{*}$$

Let $t, t_0 \in r$, where $t = \langle x, y, z \rangle$, $t_0 = \langle x_0, y_0, z_0 \rangle$. Since $z_0 \in Z(x)$, we have $x_0 \approx x$, which implies, $y \in Y(x_0)$. On the other hand $Y(x_0) \cong Y(x_0z_0)$, we have also $\exists t_1 = \langle x_1, y_1, z_1 \rangle \in r$ such that $y_1 \in Y(x_0z_0)$ and $y \approx y_1$. It means that $x_0 \approx x_1$, $z_0 \approx z_1$ and $y \approx y_1$. By transitivity of equivalence relation (\approx), we get $x \approx x_1$. Consider tuple t_1 , we found the existing of t_1 in (*) is pointed (let t_1 then t_1 in t_2 in t_3 in t_4 in t_4 in t_4 in t_5 in t_6 in t_7 in t_8 i

(A7) If $X \approx Y$ holds, then $X \approx Y$.

We need to show

$$Y(x) \cong Y(xz) \ \forall t = \langle x, y, y \rangle \in r.$$
 (**)

Let $y_0 \in Y(x)$, clearly $x_0 \approx x$. Because $X \approx Y$ is valid in r, we have $y_0 \approx y$. It is easy to see that $y \in Y(xz)$ and $y_0 \approx y$. The proof is complete.

(A8): If $X \approx \longrightarrow Y$ holds, $Z \subseteq Y$, $W \cap Y = \emptyset$, and $W \approx \longrightarrow Z$, then $X \approx \longrightarrow Z$ holds.

Assume the contrary that we have a fuzzy relation r in which $X \approx Y$ and $W \approx Z$ hold, where $Z \subseteq Y$, $W \cap Y = \emptyset$ but $X \approx Z$ does not hold.

Thus, $\exists t_1, t_2 \in r$ such that $(t_1[X] \approx t_2[X])$ is true but $(t_1[Z] \approx t_2[Z])$ is not valid. (***)

Obviously $t_2[Y] \in Y(t_1[X])$, from $t_1[X] \approx t_2[X]$. Since $X \approx \to Y$ holds then $\exists t_3 \in r : t_3[Y] \in Y(t_1[X]t_1[R-XY])$ and $t_3[Y] \approx t_2[Y]$, which implies

$$t_3[X] \approx t_1[X],\tag{1}$$

$$t_3[R - XY] \approx t_1[R - XY],\tag{2}$$

$$t_3[Y] \approx t_2[Y]. \tag{3}$$

From $W \cap Y = \emptyset$, combining with (1) and (2), we have

$$t_3[W] \approx t_1[W]. \tag{4}$$

From $Z \subseteq Y$ and (3), we have also $t_3[Z] \approx t_2[Z]$.

Since our contrary assumption (***) and transitivity of equivalence relation (\approx), it can be seen that $(t_3[Z] \approx t_1[Z])$ does not hold in r (5).

But (4) and (5) contradicts $W \approx Z$ holds in r. The proof is complete.

Proof of (A5) easy to show from definition of FMVD and properties of equivalence relation (\approx). Techniques of proof for (A6) are similar to those used in [4].

We also suppose that procedure of proof for the completeness of above inference axioms is similar to the classical case.

6. CONCLUSIONS

We have suggested the structure for representing uncertain information in the form of relational database. The models, which are given by B. P. Buckles and F. E. Petry [2] and by A. K. Mazumdar [1,6], are only special cases. Based on the concept of redundancy on a set of tuples, the definitions of fuzzy dependencies (fuzzy functional dependency and fuzzy multivalued dependency) are proposed. It is interesting to note that the set of inference rules, which is similar to classical case [7], is sound and complete as well.

In order to continue, we have already begun some studies: research for extending the relational algebra in this model, and extension of this model such that it allows the presence of null values too.

REFERENCES

- [1] Bhattacharjee T. K, Mazumdar A. K., Axiomatisation of fuzzy multivalued dependencies in a fuzzy relational data model, Fuzzy Sets and Systems 96 (1998) 343-352.
- [2] Buckles B.P and Petry E., A fuzzy representation of data for relational databases, Fuzzy Sets and Systems 7 (1980) 213-226.
- [3] Codd E.F., A relational model of data for large shared data banks, Commun. ACM 13 (6) (1970) 377-387.
- [4] Ho Thuan, Ho Cam Ha, Huynh Van Nam, Some comments about "Axiomatisation of fuzzy multivalued dependencies in a fuzzy relational data model", Journal of Computer Science and Cybernetics 16 (4) (2000) 30-33.
- [5] Petry E. and Bosc P., Fuzzy Databases Principles and Applications, Kluwer Academic Publishers, 1996.
- [6] Raju K. V. and Mazumdar A. K., Functional Dependencies and lossless join decomposition of fuzzy relational database system, ACM Trans, Database System 13 (1988) 129-1966.
- [7] Ullman J. F., Principles of Database Systems, 2nd Ed, Computer Science Press, Rockvill, MD, 1984.
- [8] Zadeh L. A., Fuzzy sets, Inform. Control 12 (1965) 338-353.
- [9] Zadeh L. A., Fuzzy sets as a basis for a theory of possibility, Fuzzy Sets and Systems 1 (1978) 3-28.

Received April 10, 2001 Revised July 2, 2001

Ho Thuan - Institute of Information Technology, NCST of Viet Nam Ho Cam Ha - The Hanoi Pedagogical Institute