λ-DECOMPOSITION OF FUZZY RELATIONAL DATABASES*

Attila Kiss
Computer Centre, Loránd Eötvös University
P. O. Box 157, Budapest 112, H–1502, Hungary

Abstract: This paper shows a natural extension of data dependencies of relational databases to fuzzy relational databases. We define the truth value of a functional dependency in a fuzzy relational database and some of its properties are shown. Then with the help of the relational algebra λ-decomposition is defined and a sufficient condition is given to decide that the dependency structure implies a lossless λ-decomposition.

Keywords: Fuzzy relational database, functional dependency, closure, relational algebra, decomposition of databases.

1. INTRODUCTION

In the practice sometimes one meets uncertain data and uncertain connections, relationships between them. Using uncertain information one has to make good decisions. Several models are proposed to build the uncertainty into the representation of the data. One of them is the fuzzy model (introduced by Zadeh [7]), which generalizes the classical set theory in the following way. Instead of classical sets, it uses fuzzy sets, which are functions from the domain to the unit interval $[0,1]$. These membership functions are natural extensions of the characteristic functions related to the classical sets. We examine fuzzy relations, which store uncertain relationships between data. In the classical relational database theory in order to design good databases (no data redundancy, no update anomalies) one has to know more additional information called functional dependencies, which say that some values determine some other values. We generalize this notion for fuzzy relations and show some propositions which are useful for designing fuzzy relational databases. Then with the tools of λ-cut we define some kind of decompositions of fuzzy relational databases.

* Research supported by Hungarian Foundation for Scientific Research Grant No 334/88

2. FUNCTIONAL DEPENDENCIES IN FUZZY RELATIONS

In the literature many kinds of fuzzy relational databases are examined. We follow the next simple definition:

Definition . Let $U = \{A_1, \ldots, A_n\}$ be the set of the attributes and each A_i is assigned to the set of possible values $DOM(A_i)$.
A function $R : \times_{i=1}^{n} DOM(A_i) \rightarrow [0,1]$ is called a fuzzy relation on $\times_{i=1}^{n} DOM(A_i)$. □

For example:

A	B	C	$R(t)$
1	1	0	1
1	1	1	0.9
0	0	0	0.8
0	1	1	0.7

and for any $t \in \times_{i=1}^{n} DOM(A_i)$ that does not occur in the table, $R(t) = 0$ holds.

In the classical database theory data dependencies play important roles. We examine now only the most important dependencies, the functional dependencies $X \rightarrow Y$ $(X, Y \subseteq U)$, which can be given by the following Horn-formula of the first order logic:

$$\phi_{X \rightarrow Y} \equiv \forall\, t_1, t_2((R(t_1) \wedge\ R(t_2) \wedge\ t_1[X] = t_2[X]) \rightarrow\ t_1[Y] = t_2[Y])$$

We get a fuzzy semantic of a formula in the usual way when substitute logic operators $\wedge$, $\forall$ with the operators inf and $\vee$, $\exists$ with sup and $\rightarrow$ is the following implication:

$$a \rightarrow b = \begin{cases} 1, & \text{if } a \leq b, \\ 1 - (a - b), & \text{otherwise} \end{cases}$$

and finally $\neg$ is

$$\neg a = 1 - a\ .$$

In this way we get the truth value of the fuzzy relation R satisfies a given functional dependency $X \rightarrow Y$:

$$T_R(X, Y) = 1 - \vee\ ((R(t_1) \wedge R(t_2))\ \mid\ t :\ t_1[X] = t_2[X]\ ,\ t_1[Y] \neq t_2[Y]).$$

For example the previous fuzzy relation generates the following truth values:
$T_R(A, B) = 0.3$, $T_R(B, C) = 0.1$, $T_R(C, A) = 0.2$, $T_R(A, C) = 0.1$, $T_R(B, A) = 0.3$, $T_R(C, B) = 0.2$, $T_R(AC, B) = 1$, $T_R(BC, A) = 0.3$, $T_R(AB, C) = 0.1$, $T_R(AB, B) = 1$, $T_R(AB, A) = 1$.

The following properties can be easily verified for the truth values, where we denote the union of X and Y by XY (See [1],[2]):
A1) If $Y \subseteq X$, then $T_R(X,Y) = 1$,
A2) $T_R(X,Y) \wedge T_R(Y,Z) \leq T_R(X,Z)$,
A3) $T_R(X,Y) \leq T_R(XZ,YZ)$.

From these other properties can be obtained:
B1) $T_R(X,Y) \wedge T_R(X,Z) \leq T_R(X,YZ)$,
B2) $T_R(X,Y) \wedge T_R(WY,Z) \leq T_R(XW,Z)$,
B3) if $Z \subseteq Y$, then $T_R(X,Y) \leq T_R(X,Z)$.

An important consequence that $T_R(X,Y) = \wedge(T_R(X,A) \mid A : A \in Y)$.

Thus a fuzzy relation generates also a fuzzy relation $T_R(X,Y)$ on U^2 with the properties A1-A3.

Moreover if there is given an arbitrary fuzzy relation $T(X,Y)$ on U^2, then it defines the fuzzy relation $T^+(X,Y)$ which is the smallest fuzzy relation on U^2 that contains $T(X,Y)$ and has the properties A1-A3. We call $T^+(X,Y)$ the closure of $T(X,Y)$. (Remind that $T_1(X,Y) \subseteq T_2(X,Y)$ iff $T_1(X,Y) \leq T_2(X,Y)$ for all $X,Y \subseteq U$).
The closure is well defined, because the fuzzy relation $S(X,Y) \equiv 1$ satisfies A1-A3 and contains every fuzzy relation on U^2 and if $T \subseteq S_1$, $T \subseteq S_1$ where S_1, S_2 satisfy A1-A3, then $T \subseteq S_1 \cap S_2$ and $S_1 \cap S_2$ also satisfies A1-A3. ($S_1 \cap S_2(X,Y) := S_1(X,Y) \wedge S_2(X,Y)$ for all $X,Y \subseteq U$.)

Proposition 1. *$T^+(X,Y)$ is a closure, that is*
1) $T(X,Y) \subseteq T^+(X,Y)$,
2) $T^{++}(X,Y) = T^+(X,Y)$,
3) if $T_1(X,Y) \subseteq T_2(X,Y)$ then $T_1^+(X,Y) \subseteq T_2^+(X,Y)$. □

Proof. These can be obtained from the definition, using that the closure is the smallest with the given properties. ■

Hence if we use T^+ for a fuzzy relation on U^2, we suppose that T^+ has the properties A1-A3.

Now we extend $T^+(X,A)$ for fuzzy sets X as follows:
Let X be a fuzzy set on U.

$$T_f^+(X,A) = \bigvee_{Z,\lambda : Z_\lambda \subseteq X} (T^+(Z,A) \wedge \lambda),$$

where for $\lambda \in [0,1]$ we define $Z_\lambda(A) = \begin{cases} \lambda, & \text{if } A \in Z \\ 0, & \text{otherwise.} \end{cases}$

We remark that for U finite there exists some $Z \subseteq U$, for which $T_f^+(X,A) = T^+(Z,A) \wedge \lambda$, where $Z_\lambda \subseteq X$ and $\lambda = \bigwedge_{A \in Z} X(A)$.

With the help of $T_f^+(X, A)$ we define a closure on U as follows:
Let X be a fuzzy set on U. Then X^+ is also a fuzzy set on U and defined by $X^+(A) = T_f^+(X, A)$ for all $A \in U$.

First we note that $T_f^+(X, A) = T^+(X, A)$ if X is a classical set, that is $X(A) = 1$ or 0 for all $A \in U$. This is true because $T^+(X, A)$ is an increasing function in the argument X.

Proposition 2. *X^+ is a closure on U, that is*
1) $X \subseteq X^+$,
2) if $X \subseteq Y$, then $X^+ \subseteq Y^+$,
3) $X^{++} = X^+$. □

Proof. See [3]. ■

We can give a simple algorithm to compute the closure of a given dependency structure $T(X, Y)$. For this we will represent $T(X, Y)$ on a fuzzy graph. (We mean that fuzzy graph is labelled directed graph, where the label-values are from the unit interval.)

Let $T(X, Y)$ be a fuzzy relation on U^2. We correspond a fuzzy graph $G_T = (V, E, l)$ to it, where the vertices $v \in V$ are the ordered pairs (X, Y) and $e \in E$ is a directed edge, if $e = ((X, Y), (X, Z))$ and l is a label function from $V \cup E$ to $[0, 1]$, such that for $v = (X, Y)$ $l(v) = T(X, Y)$ and for $e = ((X, Y), (X, Z))$ $l(e) = T(Y, Z)$.

$$\underset{(X,Y)}{\overset{T(X,Y)}{\bullet}} \quad \xrightarrow{\quad T(Y,Z) \quad} \quad \underset{(X,Z)}{\overset{T(X,Z)}{\bullet}}$$

The following algorithm gives $T^+(X, Y)$ by modifying step by step the labels of the graph:

Algorithm:

1) For all $Y \subseteq X$ let $l((X, Y)) = 1$.
2) do while Stat1 or Stat2 is true,
   (where Stat1 = there exists an edge $e = (v_1, v_2)$,
      so that $l(v_2) < l(v_1) \wedge l(e)$,
    Stat2 = there are vertices $v_1 = (X, Y)$ and $v_2 = (XZ, YZ)$
      so that $l(v_2) < l(v_1)$)
   in case Stat1 let $l(v_2) := l(v_1) \wedge l(e)$ and
      for all edges $d = ((X, Y), (X, Z))$ where $v_2 = (Y, Z)$
      let $l(d) = l(v_2)$,
   in case Stat2 let $l(v_2) := l(v_1)$ and
      for all edges $d = ((W, XZ), (W, YZ))$ where $v_2 = (XZ, YZ)$
      let $l(d) = l(v_2)$,

enddo
3) $T^+(X,Y) = l(v)$, where $v = (X,Y)$.

Proposition 3. *This algorithm is correct.* □

Proof. See [3]. ■

Since $X^+(A)$ is defined by $T^+(X,A)$, when X is a classical set on U, it can be computed by this algorithm as well.

3. LOSSLESS λ-DECOMPOSITIONS

Certainly we cannot store all of our data in only one fuzzy relation, because it would be difficult to use it for its size and the connections between data would cause redundant storage. Therefore it is worth to cut our fuzzy relation into smaller ones without loosing any information of the original fuzzy relation. This can be softened by requiring only that small information can be lost when performing decomposition. We use λ-cut of fuzzy relations:

Let R be a fuzzy relation, then for any $\lambda \in [0,1]$:

$$R_\lambda(t) := \begin{cases} R(t), & \text{if } R(t) >= \lambda \\ 0, & \text{otherwise.} \end{cases}$$

Let R^+ denote all the tuples t for which $R(t) > 0$. (So R^+ is a classical relation and it is reasonable to investigate whether R^+ satisfies a given dependency $X \to Y$ or not.) We denote all the classical relations that satisfy $X \to Y$ by $SAT(X \to Y)$. The next simple properties show the relationship between the truth values of dependencies and the λ-cut of the fuzzy relations.
Let $\neg T_R(X,Y) = 1 - T_R(X,Y)$, then $\neg T_R(X,Y) = \wedge\{\lambda \mid R^+_\lambda \in SAT(X \to Y)\}$.
Further if $\lambda_1 < \lambda_2$ then from $R^+_{\lambda_1} \in SAT(X \to Y)$ follows $R^+_{\lambda_2} \in SAT(X \to Y)$.

In order to define the decomposition of the fuzzy relations we remind the fuzzy relational algebra given by Umano in 1983 [6]. The fuzzy relational algebra contains five operators and expressions can be given by finite compositions of these operators. The operators are the following.

1) Projection
   If $X \subseteq U$ then $\Pi_X(R)(s) := \bigvee_{t:t[X]=s} R(t)$.
2) Union
   $R \cup S(t) := R(t) \vee S(t)$.
3) Selection
   If $\forall t:\ F(t) \in [0,1]$ then $\sigma_F(R)(t) := R(t) \wedge F(t)$.
4) Cartesian product
   $R \times S(uv) := R(u) \wedge S(v)$.

5) Subtraction

$$R - S(t) := 0 \vee (R(t) - S(t)).$$

Proposition 4. *The composition of λ-cutting and any of the operators 1)-4) is commutative.* □

Proof. 1) $\Pi_X(R)_\lambda(s) = 0$ iff $\Pi_X(R)(s) = \bigvee_{t[X]=s} R(t) < \lambda$ iff $t[X] = s$ implies $R(t) < \lambda$ iff $\bigvee_{t[X]=s} R_\lambda(t) = 0$ iff $\Pi_X(R_\lambda)(s) = 0$.
$\Pi_X(R)_\lambda(s) = \alpha$ iff $\Pi_X(R)(s) = \bigvee_{t[X]=s} R(t) = \alpha \geq \lambda$ iff $\bigvee_{t[X]=s} R_\lambda(t) = \alpha \geq \lambda$ iff $\Pi_X(R_\lambda)(s) = \alpha$.
2) $(R \cup S)_\lambda = 0$ iff $R(t) \vee S(t) < \lambda$ iff $R_\lambda(t) \vee S_\lambda(t) = 0$.
$(R \cup S)_\lambda = \alpha \geq \lambda$ iff $R(t) \vee S(t) \geq \lambda$ iff $R_\lambda(t) \vee S_\lambda(t) = \alpha \geq \lambda$.
3) $(R \times S)_\lambda(uv) = 0$ iff $R(u) < \lambda$ or $S(v) < \lambda$ iff $R_\lambda(u) \wedge S_\lambda(v) = 0$ iff $R_\lambda \times S_\lambda(uv) = 0$.
$(R \times S)_\lambda(uv) = \alpha \geq \lambda$ iff $R(u) \geq \lambda$ and $S(v) \geq \lambda$ iff $R_\lambda(u) \wedge S_\lambda(v) = \alpha$ iff $R_\lambda \times S_\lambda(uv) = \alpha$.
4) $\sigma_F(R)_\lambda = 0$ iff $R(t) \wedge F(t) < \lambda$ iff $F(t) = 0$ or $R(t) < \lambda$ iff $R_\lambda(t) \wedge F(t) = 0$ iff $\sigma_F(R_\lambda)(t) = 0$.
$\sigma_F(R)_\lambda = \alpha \geq \lambda$ iff $R(t) \wedge F(t) \geq \lambda$ iff $F(t) = \lambda$ and $R(t) < \lambda$ iff $R_\lambda(t) \wedge F(t) = \alpha$ iff $\sigma_F(R_\lambda)(t) = \alpha$. ■

Remark that λ-cutting and subtraction are not exchangeable operators as the following sample shows it.

If $R(t) = 0.6$, $S(t) = 0.4$ and let $\lambda = 0.5$ then $(R - S)_\lambda(t) = 0.2$ but $R_\lambda(t) - S_\lambda(t) = 0.6 - 0 = 0.6$.

The most important operator is the natural join, which can be given using the above basic operators: if R and S are fuzzy relations on X_1 and X_2 respectively, then

$$R \bowtie S(t) := R(t[X_1]) \wedge S(t[X_2])$$

for all tuples t with the attributes $X_1 \cup X_2$.

Note that $R \bowtie S = \Pi_{X_1 \cup X_2}(\sigma_F(R \times S))$ where $F(ts) = 1$ iff $t[X_1 \cap X_2] = s[X_1 \cap X_2]$ and $F(ts) = 0$ otherwise.

The lossless λ-decomposition can be defined as follows:

Definition . Let $U = \bigcup_{i=1}^{k} X_i$ and R a given fuzzy relation. R satisfies the λ-join dependency $\lambda\!-\bowtie_{i=1}^{k} X_i$ or in another words the λ-decomposition is lossless iff

$$R_\lambda = (\bowtie_{i=1}^{k} \Pi_{X_i}(R))_\lambda.$$ □

Note that it is equivalent with

$$R_\lambda = \bowtie_{i=1}^{k} \Pi_{X_i}(R_\lambda).$$

This means that the 'essential' part of the fuzzy relation, the tuples which belong to the relation on a given high level, is the natural join of its projections, that is the information of the 'essential' part of the fuzzy relation is stored in the projections, too. The problem we investigate is how to decide that each member of a given class of fuzzy relations satisfies a given λ-join dependency $\lambda - \bowtie_{i=1}^{k} X_i$. If the answer is yes, then we can store instead of the members of the class their projections to $X_1, X_2, \ldots, X_k$, that needs in general less place.

In the literature [4] it is described the case when the members R of the class fullfil that R^+ satisfies a given set of functional dependencies. More precisely if R is a fuzzy relation and $R^+ \in SAT(F)$ where F is a set of functional dependencies such kind that $F \models \bowtie_{i=1}^{k} X_i$ (which means the inclusion of sets of classical relations $SAT(F) \subseteq SAT(\bowtie_{i=1}^{k} X_i)$ then R satisfies the λ-join dependency $\lambda - \bowtie_{i=1}^{k} X_i$ where $\lambda = 0$. The classical implication problem $F \models \bowtie_{i=1}^{k} X_i$ can be solved using the well-known chase algorithm which is a special form of a theorem proving procedure applied for database dependencies.

It is natural to see such kinds of fuzzy relations whose dependency structures $T_R(X,Y)$ satisfy some condition. For example we know in advance some lower bound of $T_R(X,Y)$. Let $T(X,Y)$ be a fuzzy relation on U^2. Consider all fuzzy relations for which $T^+(X,Y) \leq T_R(X,Y)$. Denote it by $SAT(T(X,Y))$. We can see a special implication problem as whether every member of $SAT(T(X,Y))$ satisfies a given λ-lossless decomposition, $T(X,Y) \models \lambda - \bowtie_{i=1}^{k} X_i$.

For a given λ and $T(X,Y)$ define a corresponding set of functional dependencies $F(T,\lambda)$, so that $X \to Y \in F(T,\lambda)$ iff $T(X,Y) \geq \lambda$.

Proposition 5. *If* $F(T,\lambda) \models \bowtie_{i=1}^{k} X_i$ *then* $T(X,Y) \models \alpha - \bowtie_{i=1}^{k} X_i$ *where* $\alpha > 1 - \lambda$. □

Proof. If $R \in SAT(T(X,Y))$ then $T_R(X,Y) \geq T(X,Y) \geq \lambda$ if $X \to Y \in F(T,\lambda)$. Thus for $X \to Y \in F(T,\lambda)$, $1 - T_R(X,Y) \leq 1 - T(X,Y) \leq 1 - \lambda < \alpha$ which means that $R_\alpha^+ \in SAT(F(T,\lambda)) \subseteq SAT(\bowtie_{i=1}^{k} X_i)$, because of the implication $F(T,\lambda) \models \bowtie_{i=1}^{k} X_i$, so we have R satisfies $\alpha - \bowtie_{i=1}^{k} X_i$. ■

An obvious consequence is that with the chase algorithm it can be decided the implication problem for λ-lossless decomposition.

For example let $U = \{A,B,C,D,E\}$ and we deal with such fuzzy relations R for which we know in advance that $0.3 \leq T_R(B,A)$, $0.35 \leq T_R(B,E)$, $0.6 \leq T_R(E,A)$, $0.65 \leq T_R(A,D)$, $0.7 \leq T_R(AD,C)$.

If we define $T(B,A) = 0.3$, $T(B,E) = 0.35$, $T(E,A) = 0.6$, $T(A,D) = 0.65$, $T(AD,C) = 0.7$ and $\lambda = 0.6$ then $F(T,0.6) = \{E \to A, A \to D, AD \to C\}$.

Use chase algorithm [5] for the initial table

A	B	C	D	E
a		c		
a			d	e
	b			e

Identifying entries because of $E \to A$ we get

A	B	C	D	E
a		c		
a			d	e
a	b			e

Identifying entries because of $A \to D$ we get

A	B	C	D	E
a		c	d	
a			d	e
a	b		d	e

Finally, identifying entries because of $AD \to C$ we get

A	B	C	D	E
a		c	d	
a		c	d	e
a	b	c	d	e

Thus the tuple (a, b, c, d, e) appears in the table which means that $F(T, 0.6) \models AC \bowtie ADE \bowtie BE$ so all fuzzy relations $R \in SAT(T(X,Y))$ satisfy $\alpha - AC \bowtie ADE \bowtie BE$ where $\alpha > 1 - 0.6$.
For example take the following $R \in SAT(T(X,Y))$ (where N is a big positive integer):

A	B	C	D	E	$R(t)$
0	0	0	0	0	1
0	3	0	0	0	1
0	4	0	0	0	1
0	5	0	0	0	1
...	...	...	...	...	1
0	$N+2$	0	0	0	1
0	1	0	0	0	0.9
2	2	0	0	2	0.8
1	0	0	1	0	0.4
1	0	1	0	1	0.3
1	1	1	1	0	0.2

Then we have $T_R(B,A)=0.6$, $T_R(B,E)=0.7$, $T_R(E,A)=0.6$, $T_R(A,D)=0.7$, $T_R(AD,C)=0.8$.
If $\alpha=0.41$ then R_α:

A	B	C	D	E	$R_{0.41}(t)$
0	0	0	0	0	1
0	3	0	0	0	1
0	4	0	0	0	1
0	5	0	0	0	1
...	...	...	...	...	1
0	$N+2$	0	0	0	1
0	1	0	0	0	0.9
2	2	0	0	2	0.8

We get the projections $R1=\Pi_{AC}(R_{0.41}), R2=\Pi_{ADE}(R_{0.41}), R3=\Pi_{BE}(R_{0.41})$:

A	C	$R1(t)$
0	0	1
2	0	0.8

,

A	D	E	$R2(t)$
0	0	0	1
2	0	2	0.8

,

B	E	$R3(t)$
0	0	1
3	0	1
4	0	1
5	0	1
...	0	1
$N+2$	0	1
1	0	0.9
2	2	0.8

It is easy to verify that $R=R1\bowtie R2\bowtie R3$, but the size of R is $6*(N+3)$ entries and the total size of $R1, R2, R3$ is $6+8+3*(N+3)$ which is less than the size of R when $N\geq 2$. We obtained that in the case of the restriction $R\in SAT(T(X,Y))$ it is better to store the projections $\Pi_{AC}(R_{0.41}), \Pi_{ADE}(R_{0.41}), \Pi_{BE}(R_{0.41})$ which contain the same information as $R_{0.41}$ and the fuzzy relation $R^{0.41}$ defined by $R^{0.41}(t)=\begin{cases}R(t), & \text{if } R(t)<0.41\\ 0, & \text{otherwise}\end{cases}$. This way we perform a vertical and horizontal decomposition of the fuzzy relations and we have not lost any information, but we need generally less place to store them.

REFERENCES

[1] A. Kiss, The truth values of dependencies of fuzzy databases. In: *Proceedings of First Finnish-Hungarian Workshop on Computer Science*, Szeged, Aug 8-11, 1989, pp. 177-186

[2] A. Kiss, On fuzzy relational databases. In: *Proceeding of Mathematical Sciences Past and Present*, Hamburg, March 18-25, 1990, Vol 3.

[3] A. Kiss, An application of fuzzy graphs in database theory. *PU.M.A. Ser. A, Vol. 1. Budapest* (1990), No. 3-4, 337-342

[4] K.V.S.V.N. Raju, A.K. Majumdar, The study of joins in fuzzy relational databases *Fuzzy Sets and Systems* bf 21(1987), 19-34

[5] J.D. Ullman, *Principles of Database Systems.* (Computer Science Press, 1983).

[6] M. Umano, Retrieval from fuzzy database by fuzzy relational algebra, *IFAC Fuzzy Information, Knowledge Information Systems, Marseille*, 1983 pp 1-6

[7] L. A. Zadeh, Fuzzy Sets as a Basis for a Theory of Possibility *Fuzzy Sets and Systems* **1.** (1978), 3-28