Data Dependencies in Extended Possibility-Based Fuzzy Relational Databases

Z. M. Ma,[1,*] W. J. Zhang,[2] W. Y. Ma,[3] F. Mili[1]
[1] *Department of Computer Science and Engineering, Oakland University, Rochester, MI 48309*
[2] *Department of Mechanical Engineering, University of Saskatchewan, 57 Campus Drive, Saskatoon, SK S7N 5A9, Canada*
[3] *Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong SAR, China*

Based on the semantic equivalence degree the formal definitions of fuzzy functional dependencies (FFDs) and fuzzy multivalued dependencies (FMVDs) are first introduced to the fuzzy relational databases, where fuzziness of data appears in attribute values in the form of possibility attributions, as well as resemblance relations in attribute domain elements, called extended possibility-based fuzzy relational databases. A set of inference rules for FFDs and FMVDs is then proposed. It is shown that FFDs and FMVDs are consistent and the inference rules are sound and complete, just as Armstrong's axioms for classic cases. © 2002 Wiley Periodicals, Inc.

1. INTRODUCTION

In real-world applications, information is often vague or ambiguous. Therefore different kinds of incomplete information have been extensively introduced into relational databases. Incomplete information can be classified into two aspects, namely, imprecise information and uncertain information.[11,16] Intuitively, the imprecision is relevant to the content of an attribute value, and it means that a choice must be made from a given range (interval or set) of values, but we do not know exactly which one to choose at present. The uncertainty is relevant to the degree of truth of its attribute value, and it means that we can apportion some, but not all, of our belief to a given value or group of values.

*Correspondence address: Department of Computer Science and Engineering, Oakland University, Rochester, MI 48309; e-mail: zongmin_ma@yahoo.com.

INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS, VOL. 17, 321–332 (2002)
 • DOI: 10.1002/int.1088

Fuzzy values have been employed to model imprecise information in databases[5,23] since Zadeh[29] proposed the concept of fuzzy sets, and the classic relational databases have been thereby extended. Three types of fuzziness can be distinguished in fuzzy relational databases, namely, possibility distributions on attribute values,[23] similarity relations on attribute domains[5] (or proximity relations[26] and resemblances[25]), and grades of membership on tuples.[24] Corresponding to these fuzzy data representations, there exist three basic types of fuzzy relational data models,[9] namely, similarity-based relational models,[5] possibility-based relational models,[23] and fuzzy relation-based relational models.[24] Based on these three types of basic fuzzy data models, there exist some types of hybrid data models[9,21,22,25] where possibility distribution and similarity (or proximity, resemblance, and closeness) relations occur in a relation simultaneously.

Data dependencies play a crucial role in logical database design, as well as database manipulation. Some attempts have been made to represent data dependencies in fuzzy relational databases, such as fuzzy functional dependencies (FFDs)[6–8,10,17–22,24,27] and fuzzy multivalued dependencies (FMVDs).[3,20,26–28] Among these data dependencies, functional dependencies are of more interest. Being the same as classic relational databases, fuzzy functional dependencies can be used as guidelines for the design of a fuzzy relational schema that is conceptually meaningful and free of certain update anomalies. Moreover, fuzzy functional dependencies and their inferences have been applied in database security, knowledge discovery (data mining), and reasoning.[12,14] Fuzzy functional dependencies have received a lot of attention. It is necessary for the definition of fuzzy functional dependencies to compare the fuzzy values of the same attributes. Therefore, several definitions of fuzzy functional dependencies have been proposed on the basis of different semantic measures. In Ref. 24 a fuzzy relation EQUAL over $U \times U$ (U is a universe of discourse) is defined as a fuzzy measure. A fuzzy functional dependency $X \hookrightarrow Y$ holds in a fuzzy relation if for any pair of tuples, the resemblance on Y values is greater than that on X values. The proposals in Refs. 17, 18, and 19 use the same definitions of fuzzy functional dependencies, with semantic distance and semantic proximity, respectively, instead of the fuzzy relation EQUAL. Based on the possibility distribution theory, the closeness degree of fuzzy data on a domain D is introduced by Chen, et al. in Refs. 6, 7, and 8. A fuzzy functional dependency $X \hookrightarrow_{\Phi} Y$ holds in a fuzzy relation if for any pair of tuples, the closeness degree for Y values is at least that of X values, or over Φ. In Ref. 10, a fuzzy functional dependency $X \hookrightarrow_{\alpha,\beta} Y$ holds in a fuzzy relation if for any pair of tuples, the resemblance on X values is greater than the threshold α implies that the resemblance on Y values is greater than the threshold β. Regarding the issue of fuzzy functional dependencies, an overview is made by Bosc, Dubois, and Prade in Ref. 4, in which different proposals for fuzzy functional dependencies are analyzed, the connection between fuzzy functional dependencies and database design is addressed, and some semantics and use of fuzzy functional dependencies are suggested. Tripathy and Sakena express fuzzy multivalued dependencies in terms of particularization and Hamming.[28] Sozat and Yazici [27] study fuzzy functional and multivalued dependencies in similarity-based fuzzy relational database models. With semantic proximity, fuzzy functional, multivalued, and join dependencies are given in Ref. 20. Based on the fuzzy relation EQUAL[24] and its extension, fuzzy multivalued dependencies are also defined in Refs. 3 and 15, respectively.

In this article, we concentrate on the most general fuzzy relational database model, where fuzziness of data appears in attribute values in forms of possibility attributions, as well as resemblance relations in attribute domain elements, called extended possibility-based fuzzy relational databases.[4,17] Based on the semantic equivalence degree, the formal definitions of fuzzy functional dependencies (FFDs) and fuzzy multivalued dependencies (FMVDs) are first introduced to extended possibility-based fuzzy relational databases. A set of inference rules for FFDs and FMVDs is then proposed. It is shown that FFDs and FMVDs are consistent and the inference rules are sound and complete just as Armstrong's axioms for classic cases.

The remainder of this article is organized as follows. Section 2 provides background on fuzzy data, semantic measure, and fuzzy relational models. Section 3 defines fuzzy functional and multivalued dependencies. The inference rules for fuzzy data dependencies are introduced in Section 4. Section 5 concludes this article.

2. BACKGROUND

2.1. Fuzzy Set and Possibility Distribution

Fuzzy data is originally described as a fuzzy set by Zadeh.[29] Let U be a universe of discourse. Then a fuzzy value on U is characterized by a fuzzy set F in U. A membership function $\mu_F: U \rightarrow [0, 1]$ is defined for the fuzzy set F, where $\mu_F(u)$, for each $u \in U$, denotes the degree of membership of u in the fuzzy set F. Thus the fuzzy set F is described as follows:

$$F = \left\{ \frac{\mu(u_1)}{u_1}, \frac{\mu(u_2)}{u_2}, \ldots, \frac{\mu(u_n)}{u_n} \right\}$$

When $\mu_F(u)$ is viewed to be a measure of the possibility that a variable X has the value u in this approach, where X takes values in U, a fuzzy value is described by a possibility distribution π_X:[31]

$$\pi_X = \left\{ \frac{\pi_X(u_1)}{u_1}, \frac{\pi_X(u_2)}{u_2}, \ldots, \frac{\pi_X(u_n)}{u_n} \right\}$$

Here, $\pi_X(u_i)$, for $u_i \in U$, denotes the possibility that u_i is true. Let π_X and F be the possibility distribution representation and the fuzzy set representation for a fuzzy value, respectively. It is apparent that $\pi_X = F$ is true.[24]

In addition, fuzzy data is represented by similarity relations in domain elements,[30] in which the fuzziness comes from the similarity relations between two values in a universe of discourse, not from the status of an object itself. Similarity relations are thus used to describe the degree similarity of two values from the same universe of discourse. A similarity relation, Sim, on the universe of discourse, U, is a mapping, $U \times U \rightarrow [0, 1]$, such that:

(a) For $\forall x \in U$, $\mathrm{Sim}(x, x) = 1$ (reflexivity)
(b) For $\forall x, y \in U$, $\mathrm{Sim}(x, y) = \mathrm{Sim}_i(y, x)$ (symmetry)
(c) For $\forall x, y, z \in U$, $\mathrm{Sim}(x, z) \geq \max_y(\min(\mathrm{Sim}(x, y), \mathrm{Sim}(y, z)))$ (transitivity)

2.2. Fuzzy Relational Data Models

In connection with the three types of fuzzy data representations, there exist two basic extended data models for fuzzy relational databases. One of the data models is based on similarity relations,[5] or proximity relations,[26] or resemblances.[25] The other one is based on possibility distributions.[23,24] The latter can further be classified into two categories, that is, tuples associated with possibilities, and attribute values representing possibility distributions. The form of an n-tuple in each of the above-mentioned fuzzy relational models can be expressed, respectively, as:

$$t = \langle p_1, p_2, \ldots, p_i, \ldots, p_n \rangle, \quad t = \langle a_1, a_2, \ldots, a_i, \ldots, a_n, d \rangle, \quad \text{and}$$
$$t = \langle \pi_{A1}, \pi_{A2}, \ldots, \pi_{Ai}, \ldots, \pi_{An} \rangle$$

where $p_i \subseteq D_i$, with D_i being the domain of attribute A_i, for $a_i \in D_i$ and $d \in (0, 1]$, π_{Ai} the possibility distribution of attribute A_i on its domain D_i, and $\pi_{Ai}(x)$, for $x \in D_i$, the possibility that x is true.

It is clear that, based on the above-mentioned basic fuzzy relational models, there should be one type of extended fuzzy relational model[9,21,22,25] where possibility distributions and resemblance relations arise in relational databases simultaneously. In this article, we focus on such fuzzy relational databases and assume the possibility that each tuple in a fuzzy relation is 1. The resemblance relation is defined as follows.

DEFINITION 1. *A fuzzy relation r on a relational schema $R(A1, A2, \ldots, An)$ is a subset of the Cartesian product of* $\mathrm{Dom}(A1) \times \mathrm{Dom}(A2) \times \cdots \times \mathrm{Dom}(An)$, *where* $\mathrm{Dom}(Ai)$ *may be a fuzzy subset, or even a set of fuzzy subsets, and there is a resemblance relation on the* $\mathrm{Dom}(Ai)$. *A resemblance relation* Res *on* $\mathrm{Dom}(Ai)$ *is a mapping,* $\mathrm{Dom}(Ai) \times \mathrm{Dom}(Ai) \rightarrow [0, 1]$, *such that*:

(a) *For all x in* $\mathrm{Dom}(Ai)$, $\mathrm{Res}(x, x) = 1$ (*reflexivity*)
(b) *For all x, y in* $\mathrm{Dom}(Ai)$, $\mathrm{Res}(x, y) = \mathrm{Res}(y, x)$ (*symmetry*)

2.3. Semantic Measure of Fuzzy Data

The semantics of fuzzy data represented by a possibility distribution corresponds to the semantic space and the semantic relationship between two fuzzy data, and can be described by the relationship between their semantic spaces.[22] The semantic inclusion degree is then employed to measure semantic inclusion, and thus measure the semantic equivalence of the fuzzy data.

DEFINITION 2. *Let π_A and π_B be two fuzzy data, and their semantic spaces be* $\mathrm{SS}(\pi_A)$ *and* $\mathrm{SS}(\pi_B)$, *respectively. Let* $\mathrm{SID}(\pi_A, \pi_B)$ *denote the degree that π_A semantically includes π_B. Then*:

$$\mathrm{SID}(\pi_A, \pi_B) = \frac{\mathrm{SS}(\pi_B) \cap \mathrm{SS}(\pi_A)}{\mathrm{SS}(\pi_B)}$$

For two fuzzy data π_A and π_B, the meaning of $SID(\pi_A, \pi_B)$ is the percentage of the semantic space of π_B that is wholly included in the semantic space of π_A.

Following Definition *2,* the concept of equivalence degree can be easily drawn as follows.

DEFINITION 3. *Let π_A and π_B be two fuzzy data and* $\mathrm{SID}(\pi_A, \pi_B)$ *be the degree that π_A semantically includes π_B. Let* $\mathrm{SE}(\pi_A, \pi_B)$ *denote the degree that π_A and π_B are equivalent to each other. Then*:

$$\mathrm{SE}(\pi_A, \pi_B) = \min(\mathrm{SID}(\pi_A, \pi_B), \mathrm{SID}(\pi_B, \pi_A))$$

In Ref. 22, for two fuzzy data represented by possibility distributions, the definition for calculating the semantic inclusion degree of two fuzzy data is given based on possibility distributions and resemblance relations.

DEFINITION 4. *Let $U = \{u_1, u_2, \ldots, u_n\}$ be the universe of discourse. Let π_A and π_B be two fuzzy data on U based on a possibility distribution, and $\pi_A(u_i)$, for $u_i \in U$, denote the possibility that u_i is true. Let* Res *be a resemblance relation on domain U, and α for $0 \leq \alpha \leq 1$ be a threshold corresponding to* Res. $\mathrm{SID}(\pi_A, \pi_B)$ *is then defined by*

$$\mathrm{SID}(\pi_A, \pi_B) = \frac{\sum_{i,j=1}^{n} \min_{u_i, u_j \in D \text{ and } \mathrm{Res}(u_i, u_j) \geq \alpha} (\pi_B(u_i), \pi_A(u_j))}{\sum_{i=1}^{n} \pi_B(u_i)}$$

The notion of the semantic equivalence degree of attribute values can be extended to the semantic equivalence degree of tuples. Let $t_i = \langle a_{i1}, a_{i2}, \ldots, a_{in} \rangle$ and $t_j = \langle a_{j1}, a_{j2}, \ldots, a_{jn} \rangle$ be two tuples in fuzzy relational instance r over schema R $(A_1, A_2, \ldots, A_n)$. The semantic equivalence degree of tuples t_i and t_j is denoted $\mathrm{SE}(t_i, t_j) = \min\{\mathrm{SE}(t_i[A_1], t_j[A_1]), \mathrm{SE}(t_i[A_2], t_j[A_2]), \ldots, \mathrm{SE}(t_i[A_n], t_j[A_n])\}$.

PROPOSITION 1. *Let r be a fuzzy relation on schema R, U be the set of attributes of R, and $X, Y \subseteq U$. Then* $\mathrm{SE}(t[XY], s[XY]) \leq \mathrm{SE}(t[X], s[X])$ *and* $\mathrm{SE}(t[XY], s[XY]) \leq \mathrm{SE}(t[Y], s[Y])$ *for any t and s in $r(R)$.*

The proof follows directly from $\mathrm{SE}(t[XY], s[XY]) = \min\{\mathrm{SE}(t[X], s[X]), \mathrm{SE}(t[Y], s[Y])\}$. Generally, if $X \subseteq Y$, then $\mathrm{SE}(t[X], s[X]) \geq \mathrm{SE}(t[Y], s[Y])$.

PROPOSITION 2. $\mathrm{SE}(t[X], t[X]) \geq \mathrm{SE}(t[X], s[X])$ *for any t and s in a fuzzy relation $r(R)$, where $X \subseteq U$.*

Since $\mathrm{SS}(t[X]) \cap \mathrm{SS}(s[X]) \leq \mathrm{SS}(t[X])$, the proof follows directly from Definition 2 and Definition 3.

3. FUZZY DATA DEPENDENCIES

Integrity constraints play a critical role in a logical database design in which data dependencies are of more interest. One of the most important data dependencies is the functional dependency (FD) in relational databases, representing the dependency relationships among attribute values in a relation. In classic relational databases, functional dependencies can be defined as follows.

DEFINITION 5. *For a classic relation* $r(R)$, *in which* R *denotes the schema*, *its attribute set is denoted by* U, *and* $X, Y \subseteq U$, *we say* r *satisfies the functional dependency* FD: $X \rightarrow Y$, *if*:

$$(\forall t \in r)(\forall s \in r)(t[X] = s[X] \Rightarrow t[Y] = s[Y])$$

Multivalued dependencies (MVDs), originated by Fagin,[13] are another important type of data dependency that are imposed on the tuples of relational databases, relating an attribute value or a set of attribute values to a set of attribute values, independent of the other attributes in the relation. In classic relational databases, multivalued dependencies can be defined as follows.

DEFINITION 6. *For a classic relation* $r(R)$, *in which* R *denotes the schema*, *its attribute set is denoted by* U, $X, Y \subseteq U$, *and* $Z = U - XY$, *we say* r *satisfies the multivalued dependency* MVD: $X \rightarrow\rightarrow Y$, *if*:

$$(\forall t \in r)(\forall s \in r)(t[X] = s[X] \Rightarrow (\exists u \in r)(u[X] = t[X] \wedge u[Y] = t[Y] \wedge u[Z] = s[Z]))$$

3.1. Fuzzy Functional Dependencies (FFDs)

Fuzzy functional dependencies can reflexively represent the dependency relationships among attribute values in fuzzy relations, such as "the salary almost depends on the job position and experience." Following the notion of semantic equivalence degree and the methodology for the definition of fuzzy functional dependencies introduced in Ref. 22, we give the definition of fuzzy functional dependencies as follows.

DEFINITION 7. *For a relation instance* $r(R)$, *where* R *denotes the schema*, *its attribute set is denoted by* U, *and* $X, Y \subseteq U$, *we say* r *satisfies the fuzzy functional dependency* FFD: $X \hookrightarrow Y$, *if*:

$$(\forall t \in r)(\forall s \in r)(\mathrm{SE}(t[X], s[X]) \leq \mathrm{SE}(t[Y], s[Y]))$$

Consider the fuzzy relation instance r in Table I. Assume that attribute domains $Dom(X) = \{a, b, c, d, e\}$ and $Dom(Y) = \{f, g, h, i, j\}$. There are two resemblance relations $Res(X)$ and $Res(Y)$ on X and Y shown in Figure 1 and Figure 2, respectively. Let two thresholds on $Res(X)$ and $Res(Y)$ be $\alpha_1 = 0.90$ and $\alpha_2 = 0.95$, respectively.

Table I. Fuzzy relation instance r

	K	X	Y
t	1001	{0.7/a, 0.4/b, 0.5/d}	{0.9/f, 0.6/g, 1.0/h}
s	1002	{0.5/a, 0.4/c, 0.8/d}	{0.6/g, 0.9/h, 0.9/i}
u	1003	{0.3/d, 0.8/e}	{0.6/h, 0.4/i, 0.1/j}

Res(X)	a	b	c	d	e
a	1.0	0.2	0.3	0.2	0.4
b		1.0	0.92	0.4	0.1
c			1.0	0.1	0.3
d				1.0	0.2
e					1.0

Figure 1. Resemblance relation Res(X) on X.

Since $\mathrm{SE}(t[X], s[X]) = \min(\mathrm{SID}(t[X], s[X]), \mathrm{SID}(s[X], t[X])) = \min(0.824, 0.875) = 0.824$, and $\mathrm{SE}(t[Y], s[Y]) = \min(\mathrm{SID}(t[Y], s[Y]), \mathrm{SID}(s[Y], t[Y])) = \min(1.0, 0.96) = 0.96$, then $\mathrm{SE}(t[X], s[X]) \leq \mathrm{SE}(t[Y], s[Y])$ is true. Similarly, we have $\mathrm{SE}(t[X], u[X]) \leq \mathrm{SE}(t[Y], u[Y])$, and $\mathrm{SE}(s[X], u[X]) \leq \mathrm{SE}(s[Y], u[Y])$. Hence, FFD: $X \hookrightarrow Y$ holds in r.

THEOREM 1. *A classic functional dependency* FD *satisfies the definition of* FFD.

Proof. Let FD: $X \rightarrow Y$ be true. Then for $\forall t \in r$ and $\forall s \in r$, $t[X] = s[X] \Rightarrow t[Y] = s[Y]$. It is evident that $\mathrm{SE}(t[X], s[X]) = \min(\mathrm{SID}(t[X], s[X]), \mathrm{SID}(s[X], t[X])) = 1$, and $\mathrm{SE}(t[Y], s[Y]) = \min(\mathrm{SID}(t[Y], s[Y]), \mathrm{SID}(s[Y], t[Y])) = 1$. ■

3.2. Fuzzy Multivalued Dependencies (FMVDs)

Being similar to the fuzzy functional dependency, we can define the fuzzy multivalued dependencies based on the notion of the semantic equivalence degree as follows.

DEFINITION 8. *Let $r(R)$ be a fuzzy relation instance on schema R, U be the set of attributes of R, $X, Y \subseteq U$, and $Z = U - XY$. We say r satisfies the fuzzy multivalued dependency* FMVD: $X \hookrightarrow\hookrightarrow Y$ *if*:

$$(\forall t \in r)(\forall s \in r)(\exists u \in r)(\mathrm{SE}(u[X], t[X]) \geq \mathrm{SE}(t[X], s[X]) \wedge \mathrm{SE}(u[Y], t[Y]) \\ \geq \mathrm{SE}(t[X], s[X]) \wedge \mathrm{SE}(u[Z], s[Z]) \geq \mathrm{SE}(t[X], s[X]))$$

Consider the fuzzy relation instance r in Table II. Assume that attribute domains $\mathrm{Dom}(X) = \{a, b, c, d, e\}$, $\mathrm{Dom}(Y) = \{f, g, h, i, j\}$, and $\mathrm{Dom}(Z) = \{a, b, c, d, e, f\}$. There are three resemblance relations Res(X), Res(Y), and Res(Z) on X, Y, and Z in Figure 1, Figure 2, and Figure 3, respectively. Let three thresholds on Res(X), Res(Y), and Res(Z) be $\alpha 1 = 0.90$, $\alpha 2 = 0.95$, and $\alpha 3 = 0.90$, respectively.

Res(Y)	f	g	h	i	j
f	1.0	0.3	0.2	0.96	0.2
g		1.0	0.4	0.2	0.3
h			1.0	0.3	0.1
i				1.0	0.4
j					1.0

Figure 2. Resemblance relation Res(Y) on Y.

Table II. Fuzzy relation instance r

	X	Y	Z
t	{0.4/a, 0.6/b, 0.7/d}	{0.6/g, 0.9/h, 0.8/i}	{0.5/a, 0.7/c, 0.4/e}
s	{0.4/c, 0.5/d, 0.2/e}	{0.3/h, 0.6/i, 1.0/j}	{0.2/b, 0.5/c, 0.9/e, 0.8/f}
u	{0.4/a, 0.5/b, 0.6/d}	{0.6/g, 0.7/h, 0.8/f}	{0.6/d, 1.0/e, 0.7/f}

Following Definition 3 and Definition 4, we have:

$$\begin{aligned}
\mathrm{SE}(t[X], s[X]) &= \min(\mathrm{SID}(t[X], s[X]), \mathrm{SID}(s[X], t[X])) \\
&= \min(0.818, 0.529) = 0.529 \\
\mathrm{SE}(t[X], u[X]) &= \min(\mathrm{SID}(t[X], u[X]), \mathrm{SID}(u[X], t[X])) \\
&= \min(1.0, 0.882) = 0.882 > \mathrm{SE}(t[X], s[X]) \\
\mathrm{SE}(t[Y], u[Y]) &= \min(\mathrm{SID}(t[Y], u[Y]), \mathrm{SID}(u[Y], t[Y])) \\
&= \min(1.0, 0.913) = 0.913 > \mathrm{SE}(t[X], s[X]) \\
\mathrm{SE}(s[Z], u[Z]) &= \min(\mathrm{SID}(s[Z], u[Z]), \mathrm{SID}(u[Z], s[Z])) \\
&= \min(0.913, 0.875) = 0.875 > \mathrm{SE}(t[X], s[X])
\end{aligned}$$

Hence, FMVD: $X \hookrightarrow\hookrightarrow Y$ holds in r.

THEOREM 2. *A classic multivalued dependency* MVD *satisfies the definition of* FMVD.

Proof. Let relational instance $r(R)$ satisfy MVD: $X \rightarrow\rightarrow Y$, where $X, Y \subseteq R$, and $Z = R - XY$. Then $\forall t \in r, \forall s \in r$, and $t[X] = s[X] \Rightarrow (\exists u \in r)\,(u[X] = t[X] \wedge u[Y] = t[Y] \wedge u[Z] = s[Z])$. Accordingly, $\mathrm{SE}(u[X], t[X]) = \mathrm{SE}(u[Y], t[Y]) = \mathrm{SE}(u[Z], s[Z]) = \mathrm{SE}(t[X], s[X]) = 1$. ■

4. REFERENCE RULES FOR FUZZY DATA DEPENDENCIES

In classic relational databases, functional and multivalued dependencies satisfy the inference rules, namely, the axiom systems.[1,2] According to the definitions of the fuzzy functional and multivalued dependencies based on the semantic equivalence degree, a set of the inference rules for FFD and FMVD can be derived, which

Res(Z)	a	b	c	d	e	f
a	1.0	0.1	0.4	0.3	0.1	0.1
b		1.0	0.2	0.3	0.2	0.2
c			1.0	0.95	0.5	0.3
d				1.0	0.3	0.1
e					1.0	0.4
f						1.0

Figure 3. Resemblance relation Res(Z) on Z.

are similar to that for FD and MVD in classic relational databases. We call them fuzzy axiom systems. It can be proven that the fuzzy axiom systems are sound and complete.

4.1. Inference Rules for FFDs

FA1 (Reflexivity): If $Y \subseteq X \subseteq U$, then $X \hookrightarrow Y$.
FA2 (Augmentation): If $X \hookrightarrow Y$ and $Z \subseteq U$, then $XZ \hookrightarrow YZ$.
FA3 (Transitivity): If $X \hookrightarrow Y$ and $Y \hookrightarrow Z$, then $X \hookrightarrow Z$.
FA9 (Union): If $X \hookrightarrow Y$ and $X \hookrightarrow Z$, then $X \hookrightarrow YZ$.
FA10 (Decomposition): If $X \hookrightarrow YZ$, then $X \hookrightarrow Y$ and $X \hookrightarrow Z$.
FA11 (Pseudotransitivity): If $X \hookrightarrow Y$ and $YW \hookrightarrow Z$, then $XW \hookrightarrow Z$.

THEOREM 3. *The inference rules FA1–FA3 and FA9–FA11 are sound.*

Proof.

(1) Since $Y \subseteq X$, we have $\mathrm{SE}(t[X], s[X]) \leq \mathrm{SE}(t[Y], s[Y])$ for $\forall t \in r$ and $\forall s \in r$ from the definition of the semantic equivalence of tuples.
(2) Since FFD: $X \hookrightarrow Y$ holds in a relation r, $\mathrm{SE}(t[X], s[X]) \leq \mathrm{SE}(t[Y], s[Y])$ for $\forall t \in r$ and $\forall s \in r$, we have $\min(\mathrm{SE}(t[X], s[X]), \mathrm{SE}(t[Z], s[Z])) \leq \min(\mathrm{SE}(t[Y], s[Y]), \mathrm{SE}(t[Z], s[Z]))$, i.e., $\mathrm{SE}(t[XZ], s[XZ]) \leq \mathrm{SE}(t[YZ], s[YZ])$.
(3) If $X \hookrightarrow Y$ and $Y \hookrightarrow Z$, then $\mathrm{SE}(t[X], s[X]) \leq \mathrm{SE}(t[Y], s[Y])$ and $\mathrm{SE}(t[Y], s[Y]) \leq \mathrm{SE}(t[Z], s[Z])$ for $\forall t \in r$ and $\forall s \in r$, thus $\mathrm{SE}(t[X], s[X]) \leq \mathrm{SE}(t[Z], s[Z])$, that is, $X \hookrightarrow Z$.
(4) The decomposition rule and the pseudotransitivity rule follow easily from FA1–FA3.
(5) Now we prove the union rule. Since $X \hookrightarrow Y$, we may augment X to $X \hookrightarrow XY$ by FA2. Also for $X \hookrightarrow Z$, we may augment Y to $XY \hookrightarrow YZ$ by FA2. By transitivity, $X \hookrightarrow XY$ and $XY \hookrightarrow YZ$ imply $X \hookrightarrow YZ$. ∎

4.2. Inference Rules for FMVDs

FA4 (Complementation): If $X \hookrightarrow\hookrightarrow Y$, then $X \hookrightarrow\hookrightarrow (U - XY)$.
FA5 (Augmentation): If $X \hookrightarrow\hookrightarrow Y$ and $V \subseteq W$, then $WX \hookrightarrow\hookrightarrow VY$.
FA6 (Transitivity): If $X \hookrightarrow\hookrightarrow Y$ and $Y \hookrightarrow\hookrightarrow Z$, then $X \hookrightarrow\hookrightarrow (Z - Y)$.
FA12 (Union): If $X \hookrightarrow\hookrightarrow Y$ and $X \hookrightarrow\hookrightarrow Z$, then $X \hookrightarrow\hookrightarrow YZ$.
FA13 (Decomposition): If $X \hookrightarrow\hookrightarrow Y$ and $X \hookrightarrow\hookrightarrow Z$, then $X \hookrightarrow\hookrightarrow Y \cap Z$ and $X \hookrightarrow\hookrightarrow (Y - Z)$.
FA14 (Pseudotransitivity): If $X \hookrightarrow\hookrightarrow Y$ and $YW \hookrightarrow\hookrightarrow Z$, then $XW \hookrightarrow\hookrightarrow (Z - YW)$.

THEOREM 4. *The inference rules FA4–FA6 and FA12–FA14 are sound.*

Proof. Proofs of FA4, FA6, and FA12 follow directly from the proofs in Refs. 2, 3, and 28. The decomposition rule follows easily from FA4 and FA12, and the pseudotransitivity rule follows easily from FA5 and FA6.

We prove FA5 in the following. Since FMVD: $X \hookrightarrow\hookrightarrow Y$ holds in a relation r, $\mathrm{SE}(u[X], t[X]) \geq \mathrm{SE}(t[X], s[X])$, $\mathrm{SE}(u[Y], t[Y]) \geq \mathrm{SE}(t[X], s[X])$, and $\mathrm{SE}(u[U - XY], s[U - XY]) \geq \mathrm{SE}(t[X], s[X])$ for t, s, and $u \in r$. We have $\min(\mathrm{SE}(u[X], t[X]), \mathrm{SE}(u[W], t[W])) \geq \min(\mathrm{SE}(t[X], s[X]), (t[W], s[W]))$ and

$\min(\mathrm{SE}(u[Y], t[Y]), \mathrm{SE}(u[V], t[V])) \geq \min(\mathrm{SE}(t[X], s[X]), \mathrm{SE}(u[V], t[V])) \geq \min(\mathrm{SE}(t[X], s[X]), (t[W], s[W]))$ because of $V \subseteq W$. Besides, $\mathrm{SE}(u[U - XYWV], s[U - XYWV]) \geq \mathrm{SE}(u[U - XY], s[\underline{U}-XY]) \geq \mathrm{SE}(t[X], s[X]) \geq \mathrm{SE}(t[WX], s[WX])$. Thus, $WX \hookrightarrow\hookrightarrow VY$ holds in r. ■

4.3. Mixed Inference Rules for FFDs and FMVDs

FA7: If $X \hookrightarrow Y$, then $X \hookrightarrow\hookrightarrow Y$.
FA8: If $X \hookrightarrow\hookrightarrow Y$, $Z \subseteq Y$, $W \cap Y = \Phi$, and $W \hookrightarrow Z$, then $X \hookrightarrow Z$.

THEOREM 5. *The inference rules FA7–FA8 are sound.*

Proof.

(1) Suppose $X \hookrightarrow\hookrightarrow Y$ does not hold in r. Then $\mathrm{SE}(u[Y], t[Y]) < \mathrm{SE}(t[X], s[X])$ or $\mathrm{SE}(u[Z], s[Z]) < \mathrm{SE}(t[X], s[X])$ when $\mathrm{SE}(u[X], t[X]) \geq \mathrm{SE}(t[X], s[X])$. Since $X \hookrightarrow Y$ holds in r, we have $\mathrm{SE}(t[X], s[X]) \leq \mathrm{SE}(t[Y], s[Y])$, $\mathrm{SE}(u[X], t[X]) \leq \mathrm{SE}(u[Y], t[Y])$, and $\mathrm{SE}(u[X], s[X]) \leq \mathrm{SE}(u[Y], s[Y])$. If $\mathrm{SE}(u[Y], t[Y]) < \mathrm{SE}(t[X], s[X])$, then we have $\mathrm{SE}(t[X], s[X]) \leq (u[X], t[X]) \leq \mathrm{SE}(u[Y], t[Y]) < \mathrm{SE}(t[X], s[X])$. There is a contradiction. Similarly, using FA2 (Augmentation) and Proposition 1, we can draw the conclusion that there exists a contradiction if $\mathrm{SE}(u[Z], s[Z]) < \mathrm{SE}(t[X], s[X])$.

(2) Suppose that $X \hookrightarrow\hookrightarrow Y$ and $W \hookrightarrow Z$ hold in a relation $r(R)$, where $W \cap Y = \Phi$ and $Z \subseteq Y$, but $X \hookrightarrow Z$ does not hold in r. Then there are tuples t and s in r such that $\mathrm{SE}(t[X], s[X]) > \mathrm{SE}(t[Z], s[Z])$. By $X \hookrightarrow\hookrightarrow Y$ applied to t and s, there is a tuple u in r such that $\mathrm{SE}(u[X], t[X]) \geq \alpha$, $\mathrm{SE}(u[Y], t[Y]) \geq \alpha$, and $\mathrm{SE}(u[R - XY], s[R - XY]) \geq \alpha$, where $\mathrm{SE}(t[X], s[X]) = \alpha$. Since $W \cap Y = \Phi$ and $W \subseteq X(R - XY)$, hence $\mathrm{SE}(u[W], s[W]) \geq \mathrm{SE}(u[X(R - XY)], s[X(R - XY)]) = \min(\mathrm{SE}(u[X], s[X]), \mathrm{SE}(u[R - XY], s[R - XY])) \geq \alpha$. As $Z \subseteq Y$, $\mathrm{SE}(u[Z], t[Z]) \geq \mathrm{SE}(u[Y], t[Y]) \geq \alpha$. Hence, $\mathrm{SE}(u[Z], s[Z]) < \mathrm{SE}(t[X], s[X])$, and then $\mathrm{SE}(u[W], s[W]) > \mathrm{SE}(u[Z], s[Z])$. This contradicts $W \hookrightarrow Z$. So we can conclude that $X \hookrightarrow Z$ holds in r. ■

THEOREM 6. *The inference rules FA1–FA14 are complete.*

Proof.[20] Let F and G be the sets of FFDs and FMVDs on the universe of discourse U, respectively. The theorem means that any FFD: $f = A \hookrightarrow B$ and FMVD: $g = C \hookrightarrow\hookrightarrow D$, which are logically implied by F and G, can be deduced from F and G by FA1–FA14.

Let $(F, G)+$ be the closures of F and G. For a given FFD: $f = A \hookrightarrow B$ or FMVD: $g = C \hookrightarrow\hookrightarrow D$ that does not belong to $(F, G)+$, there exists an instance r on U such that all dependencies in F and G are valid in r, but $A \hookrightarrow B$ or $C \hookrightarrow\hookrightarrow D$ is invalid in r.

Let F' and G' be two sets of classic dependencies correspond to F and G, namely, $F' = \{X \rightarrow Y \mid X \hookrightarrow Y \in F\}$ and $G' = \{X \rightarrow\rightarrow Y \mid X \hookrightarrow\hookrightarrow Y \in G\}$. Let FD: $f' = A \rightarrow B$ and MVD: $g' = C \rightarrow\rightarrow D$. We can construct a relational instance r' that satisfies F' and G' but does not satisfy f' and g'. By Theorem 1 and Theorem 2, we know that r' satisfies F and G but does not satisfy f and g. This problem transformed into the classic correspondence problem.[2] ■

THEOREM 7. *The inference rules FA1–FA14 are sound and complete.*

The soundness of the inference rules follows from Theorem 4 and Theorem 5, and the completeness of inference rules follows from Theorem 6.

5. CONCLUSION

Fuzzy data may emerge in databases due to information imprecision or uncertainty. Fuzzy values have been used to represent imprecise data in relational databases. Data dependencies not only play a critical role in a logical database design, but also have significant influence on fuzzy data processing such as queries and updates.

Based on the most general fuzzy relational databases, where fuzziness of data appears in attribute values in the form of possibility attributions, as well as resemblance relations in attribute domain elements, called extended possibility-based fuzzy relational databases, we introduce the notion of semantic equivalence degree to evaluate semantic relationships between fuzzy data and give the expressions. With this proposal, we discuss the issues of fuzzy data dependencies. We give a set of inference rules for fuzzy functional and multivalued dependencies, and show that the inference rules for FFDs and FMVDs, which are similar to the classic case of Armstrong's axioms, are sound and complete.

Acknowledgments

The authors thank NSERC and AECL, Canada, for support to this study.

References

1. Armstrong WW. Dependency structures of data base relationships. In: Proc IFIP Congress, 1974. pp 580–583.
2. Beeri C, Fagin R, Howard JH. A complete axiomatization for functional and multivalued dependencies in database relations. In: ACM SIGMOD Conference, 1977. pp 47–61.
3. Bhattacharjee TK, Mazumdar AK. Axiomatisation of fuzzy multivalued dependencies in a fuzzy relational data model. Fuzzy Sets Syst 1998;96(3):343–352.
4. Bosc P, Dubois D, Prade H. Fuzzy functional dependencies and redundancy elimination. J American Society for Information Science 1998;49(3):217–235.
5. Buckles BP, Petry FE. A fuzzy representation of data for relational database. Fuzzy Sets Syst 1982;7(3):213–226.
6. Chen GQ. Fuzzy functional dependency as a sort of semantic knowledge: Representation, preservation and use. In Pons O, Vila MA, Kacprzyk J, editors. Knowledge management in fuzzy databases. Physica-Verlag (Springer); 2000. pp 122–133.
7. Chen GQ, Kerre EE, Vandenbulcke J. A computational algorithm for the FFD closure and a complete axiomatization of fuzzy functional dependency (FFD). Int J Intell Syst 1994;9: 421–439.
8. Chen GQ, Kerre EE, Vandenbulcke J. Normalization based on fuzzy functional dependency in a fuzzy relational data model. Inf Syst 1996;21(3):299–310.
9. Chen GQ, Vandenbulcke J, Kerre EE. A general treatment of data duplicate in a fuzzy relational data model. J American Society of Information Science 1992;43(4):304–311.
10. Cubero JC, Vila MA. A new definition of fuzzy functional dependency in fuzzy relational databases. Int J Intell Syst 1994;9(5):441–448.
11. Dubois D, Prade H. Possibility theory: An approach to computerized processing. New York: Plenum Press; 1986.
12. Dutta S. Approximate reasoning by analogy to answer null query. Int J Approximate Reasoning 1991;5:373–398.
13. Fagin R. Multivalued dependencies and a new normal form for relational databases. ACM Trans Database Syst 1977;2(3):262–278.

14. Hale J, Shenoi S. Analyzing FD inference in relational databases. Data & Knowledge Engineering 1996;18(2):167–183.
15. Jyothi S, Babu MS. Multivalued dependencies in fuzzy relational databases and lossless join decomposition. Fuzzy Sets Syst 1997;88(3):315–332.
16. Klir GJ, Folger TA. Fuzzy sets, uncertainty, and information. Englewood Cliffs, NJ: Prentice Hall; 1988.
17. Liao SY, Wang HQ, Liu WY. Functional dependencies with null values, fuzzy values, and crisp values. IEEE Trans Fuzzy Syst 1999;7(1):97–103.
18. Liu WY. The reduction of the fuzzy data domain and fuzzy consistent join. Fuzzy Sets Syst 1992;51(1):89–96.
19. Liu WY. Extending the relational model to deal with fuzzy values. Fuzzy Sets Syst 1993;60(2):207–212.
20. Liu WY. Fuzzy data dependencies and implication of fuzzy data dependencies. Fuzzy Sets Syst 1997;92(3):341–348.
21. Ma ZM, Zhang WJ, Ma WY. Functional dependencies in extended possibility-based fuzzy relational databases. In: Proc 9th IEEE Int Conf on Fuzzy System; Vol. 2, 2000. pp 929–932.
22. Ma ZM, Zhang WJ, Ma WY. Semantic measure of fuzzy data in extended possibility-based fuzzy relational databases. Int J Intell Syst 2000;15(8):705–716.
23. Prade H, Testemale C. Generalizing database relational algebra for the treatment of incomplete or uncertain information. Inf Sci 1984;34:115–143.
24. Raju KVSVN, Majumdar AK. Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database system. ACM Trans Database Syst 1988;13(2):129–166.
25. Rundensteiner EA, Hawkes LW, Bandler W. On nearness measures in fuzzy relational data models. Int J Approximate Reasoning 1989;3:267–298.
26. Shenoi S, Melton A. Proximity relations in the fuzzy relational databases. Fuzzy Sets Syst 1989;31(3):285–296.
27. Sozat MI, Yazici A. A complete axiomatization for fuzzy functional and multivalued dependencies in fuzzy database relations. Fuzzy Sets Syst 2001;117(2):161–181.
28. Tripathy RC, Sakena PC. Multivalued dependencies in fuzzy relational databases. Fuzzy Sets Syst 1990;38(3):267–279.
29. Zadeh LA. Fuzzy sets. Inf Contr 1965;8(3):338–353.
30. Zadeh LA. Similarity relations and fuzzy orderings. Inf Sci 1970;3(2):177–200.
31. Zadeh LA. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets Syst 1978;1(1):3–28.