Functional Dependencies of Data Tables Over Domains with Similarity Relations*

Radim Bělohlávek and Vilém Vychodil

Department of Computer Science, Palacky University, Olomouc
Tomkova 40, CZ-779 00 Olomouc, Czech Republic
{radim.belohlavek, vilem.vychodil}@upol.cz

Abstract. We study attribute dependencies in a setting of fuzzy logic. Our dependencies are described by formulas $A \Rightarrow B$ where A and B are fuzzy sets of attributes. The meaning of $A \Rightarrow B$ is that any two objects that have similar values on attributes from A have also similar values on attributes from B. Our approach generalizes classical functional dependencies from databases and also several approaches to functional dependencies in fuzzy databases. We show a connection to other interpretation of formulas $A \Rightarrow B$ which we studied before. Although the two interpretations are different, they have the same concept of semantic entailment. This connection enables us to find a complete axiomatization of reasoning with our attribute dependencies. We prove completeness theorem in the usual as well as in a so-called graded style. The connection also enables us to compute a complete non-redundant basis of all functional dependencies which are true in a data table. We present examples and and outline future research.

1 Introduction

Attribute dependencies of the form $A \Rightarrow B$ where A and B are collections of attributes have been studied for a long time in computer science. They play a crucial role in databases where they are called functional dependencies, see [28] for a good overview. The interpretation of a functional dependence $A \Rightarrow B$ is the following: any two objects which have the same values of attributes from A have also the same values of attributes from B. Another area is data mining where formulas $A \Rightarrow B$ are called association rules, see e.g. [37], or attribute implications in formal concept analysis, see [18, 22]. Here, formulas $A \Rightarrow B$ have the following meaning: If an object has all attributes from A then it has all attributes from B. The aim in data mining is to extract "all interesting rules" from data.

Both functional dependencies and association rules (attribute implications) have been thoroughly studied. In a paper by 29 leading scientists in databases

* Supported by grant No. 1ET101370417 of GA AV ČR, by grant No. 201/05/0079 of the Czech Science Foundation, and by institutional support, research plan MSM 6198959214.

Bhanu Prasad (Editor): IICAI-05, pp. 2486-2504, 2005.

and data representation [1], it has been pointed out that one of the important future topics in data representation is reasoning about uncertain data. In particular, one has to extend existing tools to allow for imprecision. For instance, not only exact matches but also approximate matches of data items need to be taken into account in the very foundations of data processing. From this point of view, it seems necessary to extend the notion and interpretation of classical functional dependencies so as to include similarity in attribute values. A natural idea is to interpret a functional dependence $A \Rightarrow B$ as follows: any two objects which have similar values of attributes from A have also similar values of attributes from B. For instance, "similar age and similar income imply similar savings" is a dependence which makes much more sense than its classical counterpart "the same age and the same income imply the same savings". A feasible option to put this idea into work is offered by fuzzy logic [26]. Suppose a domain D_y (i.e., the set of all values) of each attribute y is equipped with a fuzzy similarity $\approx_y$ (a particular fuzzy relation assigning to any values $a, b \in D_y$ a degree $a \approx_y b \in [0,1]$ to which a is similar to b). Then one may consider formulas $A \Rightarrow B$ with A and B being fuzzy sets of attributes, and the following meaning of $A \Rightarrow B$: for any two objects x_1, x_2, if the degree $x_1[y] \approx_y x_2[y]$ of similarity of their y-values $x_1[y], x_2[y] \in D_y$ is at least $A(y)$ for each attribute y, then for each attribute y' the degree $x_1[y'] \approx_{y'} x_2[y']$ is at least $B(y')$. Therefore, degrees $A(y) \in [0,1]$ and $B(y) \in [0,1]$ act as thresholds for similarities in attribute values. It is easily seen that this approach extends the classical one. Namely, if A and B are crisp sets (i.e. $A(y) \in \{0,1\}$ and $B(y) \in \{0,1\}$ for each $y \in Y$) and each $\approx_y$ is an ordinary equality then the above meaning coincides with the meaning of attribute dependencies.

The present paper elaborates on the above-sketched conception of functional dependencies. Our approach extends previous approaches to functional dependencies in a fuzzy setting, see [35] for an overview. Most importantly, we study semantic entailment of functional dependencies, show a connection to object-attribute semantics of formulas $A \Rightarrow B$, and present a system of deduction rules for reasoning with functional dependencies which is complete in the usual as well as in the graded style.

2 Preliminaries

Fuzzy logic and fuzzy set theory are formal frameworks for a manipulation of a particular form of imperfection called fuzziness (vagueness). Contrary to classical logic, fuzzy logic uses a scale L of truth degrees, a most common choice being $L = [0,1]$ (real unit interval) or some subchain of $[0,1]$. This enables to consider intermediate truth degrees of propositions, e.g. "object x has attribute y" has a truth degree 0.8 indicating that the proposition is almost true. In addition to a set L of truth degrees, one has to pick an appropriate collection of logical connectives (implication, conjunction, ...). A general choice of a set of truth degrees plus logical connectives is represented by so-called complete residuated lattices (equipped possibly with additional operations). The rest of this section presents in introduction to fuzzy logic notions we need in the sequel. Details can be found e.g. in [4, 21, 23], a good introduction to fuzzy logic and fuzzy sets is presented in [26].

A complete residuated lattice with a truth-stressing hedge (shortly, a hedge) [23, 24] is an algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, {}^*, 0, 1\rangle$ such that $\langle L, \wedge, \vee, 0, 1\rangle$ is a complete lattice with 0 and 1 being the least and greatest element of L, respectively; $\langle L, \otimes, 1\rangle$ is a commutative monoid (i.e. $\otimes$ is commutative, associative, and $a \otimes 1 = 1 \otimes a = a$ for each $a \in L$); $\otimes$ and $\rightarrow$ satisfy so-called adjointness property:

$$a \otimes b \leq c \quad \text{iff} \quad a \leq b \rightarrow c \tag{1}$$

for each $a, b, c \in L$; hedge * satisfies

$$1^* = 1, \tag{2}$$
$$a^* \leq a, \tag{3}$$
$$(a \rightarrow b)^* \leq a^* \rightarrow b^*, \tag{4}$$
$$a^{**} = a^*, \tag{5}$$

for each $a, b \in L$, $a_i \in L$ $(i \in I)$. Elements a of L are called truth degrees. $\otimes$ and $\rightarrow$ are (truth functions of) "fuzzy conjunction" and "fuzzy implication". Hedge * is a (truth function of) logical connective "very true", see [23, 24]. Properties (3)–(5) have natural interpretations, e.g. (3) can be read: "if a is very true, then a is true", (4) can be read: "if $a \rightarrow b$ is very true and if a is very true, then b is very true", etc.

A common choice of $\mathbf{L}$ is a structure with $L = [0, 1]$ (unit interval), $\wedge$ and $\vee$ being minimum and maximum, $\otimes$ being a left-continuous t-norm with the corresponding $\rightarrow$. Three most important pairs of adjoint operations on the unit interval are:

$$\text{Łukasiewicz:} \quad \begin{aligned} a \otimes b &= \max(a + b - 1, 0), \\ a \rightarrow b &= \min(1 - a + b, 1), \end{aligned} \tag{6}$$

$$\text{Gödel:} \quad \begin{aligned} a \otimes b &= \min(a, b), \\ a \rightarrow b &= \begin{cases} 1 \text{ if } a \leq b, \\ b \text{ otherwise}, \end{cases} \end{aligned} \tag{7}$$

$$\text{Goguen (product):} \quad \begin{aligned} a \otimes b &= a \cdot b, \\ a \rightarrow b &= \begin{cases} 1 \ \text{ if } a \leq b, \\ \frac{b}{a} \text{ otherwise}. \end{cases} \end{aligned} \tag{8}$$

In applications, we usually need a finite linearly ordered $\mathbf{L}$. For instance, one can put $L = \{a_0 = 0, a_1, \ldots, a_n = 1\} \subseteq [0, 1]$ $(a_0 < \cdots < a_n)$ with $\otimes$ given by $a_k \otimes a_l = a_{\max(k+l-n,0)}$ and the corresponding $\rightarrow$ given by $a_k \rightarrow a_l = a_{\min(n-k+l,n)}$. Such an $\mathbf{L}$ is called a finite Łukasiewicz chain. Another possibility is a finite Gödel chain which consists of L and restrictions of Gödel operations on $[0, 1]$ to L.

Two boundary cases of (truth-stressing) hedges are (i) identity, i.e. $a^* = a$ ($a \in L$); (ii) globalization [34]:

$$a^* = \begin{cases} 1 & \text{if } a = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

A special case of a complete residuated lattice with hedge is the two-element Boolean algebra $\langle \{0,1\}, \wedge, \vee, \otimes, \rightarrow, {}^*, 0, 1 \rangle$, denoted by **2**, which is the structure of truth degrees of the classical logic. That is, the operations $\wedge, \vee, \otimes, \rightarrow$ of **2** are the truth functions (interpretations) of the corresponding logical connectives of the classical logic and $0^* = 0$, $1^* = 1$. Note that if we prove an assertion for general **L**, then, in particular, we obtain a "crisp version" of this assertion for **L** being **2**.

Having **L**, we define usual notions: an **L**-set (fuzzy set) A in universe U is a mapping $A\colon U \rightarrow L$, $A(u)$ being interpreted as "the degree to which u belongs to A". If $U = \{u_1, \ldots, u_n\}$ then A can be denoted by $A = \{{}^{a_1}/u_1, \ldots, {}^{a_n}/u_n\}$ meaning that $A(u_i)$ equals a_i for each $i = 1, \ldots, n$. For brevity, we introduce the following convention: we write $\{\ldots, u, \ldots\}$ instead of $\{\ldots, {}^1/u, \ldots\}$, and we also omit elements of U whose membership degree is zero. For example, we write $\{u, {}^{0.5}/v\}$ instead of $\{{}^1/u, {}^{0.5}/v, {}^0/w\}$, etc. Let $\mathbf{L}^U$ denote the collection of all **L**-sets in U. The operations with **L**-sets are defined componentwise. For instance, the intersection of **L**-sets $A, B \in \mathbf{L}^U$ is an **L**-set $A \cap B$ in U such that $(A \cap B)(u) = A(u) \wedge B(u)$ for each $u \in U$, etc. Binary **L**-relations (binary fuzzy relations) between X and Y can be thought of as **L**-sets in the universe $X \times Y$. That is, a binary **L**-relation $I \in \mathbf{L}^{X \times Y}$ between a set X and a set Y is a mapping assigning to each $x \in X$ and each $y \in Y$ a truth degree $I(x, y) \in L$ (a degree to which x and y are related by I). An **L**-set $A \in \mathbf{L}^X$ is called crisp if $A(x) \in \{0, 1\}$ for each $x \in X$. Crisp **L**-sets can be identified with ordinary sets. For a crisp A, we also write $x \in A$ for $A(x) = 1$ and $x \notin A$ for $A(x) = 0$. An **L**-set $A \in \mathbf{L}^X$ is called empty (denoted by $\emptyset$) if $A(x) = 0$ for each $x \in X$. For $a \in L$ and $A \in \mathbf{L}^X$, $a \otimes A \in \mathbf{L}^X$ is defined by $(a \otimes A)(x) = a \otimes A(x)$.

Given $A, B \in \mathbf{L}^U$, we define a subsethood degree

$$S(A, B) = \textstyle\bigwedge_{u \in U} \bigl(A(u) \rightarrow B(u)\bigr), \tag{10}$$

which generalizes the classical subsethood relation $\subseteq$. $S(A, B)$ represents a degree to which A is a subset of B. In particular, we write $A \subseteq B$ iff $S(A, B) = 1$. As a consequence, $A \subseteq B$ iff $A(u) \leq B(u)$ for each $u \in U$.

A binary **L**-relation $\approx$ in U (i.e., between U and U) is called reflexive if for each $u \in U$ we have $u \approx u = 1$; symmetric if for each $u, v \in U$ we have $u \approx v = v \approx u$; transitive if for each $u, v, w \in U$ we have $(u \approx v) \otimes (v \approx w) \leq (u \approx w)$; **L**-equivalence if it is reflexive, symmetric, and transitive; **L**-equality if it is an **L**-equivalence for which $u \approx v = 1$ iff $u = v$.

In the following we use well-known properties of residuated lattices and fuzzy structures which can be found in monographs [4, 23]. Throughout the rest of the paper, **L** denotes an arbitrary complete residuated lattice with a hedge.

3 Functional dependencies over domains with similarity relations

3.1 Functional dependencies and their validity

First, we introduce functional dependencies in our setting. Suppose Y is a finite set of attributes. A *(fuzzy) functional denepdence (over attributes Y)* is an expression $A \Rightarrow B$, where $A, B \in \mathbf{L}^Y$ (A and B are fuzzy sets of attributes).

To get to the interpretation (meaning) of functional dependencies, we need to introduce the following concept. A *data table over domains with similarity relations* is a tuple $\mathcal{D} = \langle X, Y, \{\langle D_y, \approx_y\rangle \mid y \in Y\}, T\rangle$ where

- X is a non-empty set (of objects, table items),
- Y is a non-empty finite set (of attributes, the same as in the definition of functional dependencies),
- for each $y \in Y$, D_y is a non-empty set (of values of attribute y) and $\approx_y$ is a binary fuzzy relation which is reflexive and symmetric,
- T is a mapping assigning to each $x \in X$ and $y \in Y$ a value $T(x, y) \in D_y$ (value of attribute y on object x).

$\mathcal{D}$ will always denote some data table over domains with similarity relations with its components denoted as above.

Remark 1. (1) We will sometimes require $\approx_y$'s to be **L**-equivalences or even **L**-equalities.

(2) Consider $L = \{0, 1\}$ (case of classical logic). If each $\approx_y$ is an equality (i.e. $a \approx_y b = 1$ iff $a = b$), then $\mathcal{D}$ can be identified with what is called a relation on relation scheme Y with domains D_y ($y \in Y$) [28].

(3) For $x \in X$ and $Z \subseteq Y$, $x[Z]$ denotes a tuple of values $T(x, y)$ for $y \in Z$. We may assume that attribute from Y are numbered, i.e. $Y = \{y_1, \ldots, y_n\}$, and thus linearly ordered by this numbering, and assume that attributes in $x[Z]$ are ordered in this way. Particularly, $x[y]$ is $x[\{y\}]$ which can be identified with $T(x, y)$.

(4) $\mathcal{D}$ can be seen as a table with rows and columns corresponding to $x \in X$ and $y \in Y$, respectively, and with table entries containing values $T(x, y) \in D_y$. Moreover, each domain D_y is equipped with an additional information about similarity of elements from D_y.

Given a data table $\mathcal{D} = \langle X, Y, \{\langle D_y, \approx_y\rangle \mid y \in Y\}, T\rangle$, we want to introduce a condition for a functional dependence $A \Rightarrow B$ to be true in $\mathcal{D}$ which says basically the following: "for any two objects $x_1, x_2 \in X$: if x_1 and x_2 have similar values on attributes from A then x_1 and x_2 have similar values on attributes from A". Define first for a given $\mathcal{D}$, objects $x_1, x_2 \in X$, and a fuzzy set $C \in \mathbf{L}^Y$ of attributes a *degree $x_1(C) \approx x_2(C)$ to which x_1 and x_2 have similar values on attributes from C* (agree on attributes from C) by

$$x_1(C) \approx x_2(C) \;=\; \bigwedge_{y \in Y} \big(C(y) \to (x_1[y] \approx_y x_2[y])\big). \tag{11}$$

That is, $x_1(C) \approx x_2(C)$ is truth degree of proposition "for each attribute $y \in Y$: if y belongs to C then the value $x_1[y]$ of x_1 on y is similar to the value $x_2[y]$ of x_2 on y", which can be seen as a degree to which x_1 and x_2 have similar values on attributes from C. Then, the above idea of validity of a functional dependence is then captured by the following definition. A *degree* $||A \Rightarrow B||_{\mathcal{D}}$ *to which* $A \Rightarrow B$ *is true in* $\mathcal{D}$ is defined by

$$||A \Rightarrow B||_{\mathcal{D}} = \bigwedge_{x_1,x_2 \in X} \big((x_1(A) \approx x_2(A))^* \rightarrow (x_1(B) \approx x_2(B))\big). \qquad (12)$$

Remark 2. (1) It is easily seen that $||A \Rightarrow B||_{\mathcal{D}}$ is a truth degree of proposition "for any objects $x_1, x_2 \in X$: if it is (very) true that x_1 and x_2 have similar values on attributes from A then x_1 and x_2 have similar values on attributes from B".

(2) If A and B are crisp sets (i.e. $A(y) \in \{0,1\}$ and $B(y) \in \{0,1\}$ for each $y \in Y$) then A and B may be considered as ordinary sets and $A \Rightarrow B$ may be seen as an ordinary functional dependence. Then, if $\approx_y$ is a crisp equality (i.e., $a \approx_y b = 1$ iff $a = b$ and $a \approx_y b = 0$ iff $a \neq b$), $x_1(A) \approx x_2(A) = 1$ iff $x_1[A] = x_2[A]$ and similarly for B. Therefore, $||A \Rightarrow B||_{\mathcal{D}} = 1$ iff $A \Rightarrow B$ is true in $\mathcal{D}$ in the usual sense of validity of ordinary functional dependencies.

(3) For a functional dependence $A \Rightarrow B$, degrees $A(y) \in L$ and $B(y) \in L$ can be seen as thresholds. This is best seen when * is globalization, i.e. $1^* = 1$ and $a^* = 0$ for $a < 1$. Since for $a, b \in L$ we have $a \leq b$ iff $a \rightarrow b = 1$, we have

$$(a \rightarrow b)^* = \begin{cases} 1 & \text{iff } a \leq b, \\ 0 & \text{iff } a \not\leq b. \end{cases}$$

Therefore, $||A \Rightarrow B||_{\mathcal{D}} = 1$ means that a proposition "for any objects $x_1, x_2 \in X$: if for each attribute $y \in Y$, $A(y) \leq (x_1[y] \approx_y x_2[y])$, then for each attribute $y' \in Y$, $B(y') \leq (x_1[y'] \approx_y x_2[y'])$". As a particular example, if $A(y) = a$ for $y \in Y_A \subseteq Y$ (and $A(y) = 0$ for $y \notin Y_A$) $B(y) = b$ for $y \in Y_B \subseteq Y$ (and $B(y) = 0$ for $y \notin Y_B$), the proposition becomes "for any objects $x_1, x_2 \in X$: if for each attribute $y \in Y_A$, $x_1[y]$ is similar to $x_2[y]$ in degree at least a, then for each attribute $y' \in Y_B$, $x_1[y']$ is similar to $x_2[y']$ in degree at least b". That is, having A and B fuzzy sets allows for a rich expressibility of relationships between attributes which is why we want A and B to be fuzzy sets in general.

Remark 3. In the literature on fuzzy functional dependencies (see [33, 35] and the references therein, [35] itself provides a good overview), fuzzy functional dependencies are considered as formulas $A \Rightarrow B$ with A and B being ordinary sets. This does not allow to express relationships using thresholds as shown above. The only exception seems to be [16] where thresholds are present but are the same for A and the same for B. Due to a limited scope, we not comment on the existing approaches and postpone a discussion to a forthcoming paper. We only mention that our approach is more general than the close existing approaches in that it allows for thresholds and is not limited to truth degrees from $[0, 1]$. Furthermore, we provide more theoretical insight.

3.2 Relationships to fuzzy attribute implications: transformation theorems

In this section, we are going to show an important connection between our functional dependencies and so-called fuzzy attribute implications. Fuzzy attribute implications were introduced in [31]. [7, 9, 11–13] contain results we will refer to. A (*fuzzy*) *attribute implication* (over attributes from Y) is just another term for (fuzzy) functional dependence, i.e. a (fuzzy) attribute implication is an expression $A \Rightarrow B$ where $A, B \in \mathbf{L}^Y$ are fuzzy sets of attributes. Fuzzy attribute dependencies are interpreted in so-called data tables with fuzzy attributes. A *data table with fuzzy attributes* can be seen as a triplet $\mathcal{T} = \langle X, Y, I\rangle$ where X is a set of objects, Y is a finite set of attributes (the same as above in the definition of a fuzzy attribute implication), and $I \in \mathbf{L}^{X\times Y}$ is a binary $\mathbf{L}$-relation between X and Y assigning to each object $x \in X$ and each attribute $y \in Y$ a degree $I(x, y)$ to which x has y. $\mathcal{T} = \langle X, Y, I\rangle$ can be thought as a table with rows and columns corresponding to objects $x \in X$ and attributes $y \in Y$, respectively, and table entries containing degrees $I(x, y)$. A row of a table $\mathcal{T} = \langle X, Y, I\rangle$ corresponding to an object $x \in X$ can be seen as a fuzzy set of attributes to which an attribute $y \in Y$ belongs to a degree $I(x, y)$, i.e. a fuzzy set $I_x \in \mathbf{L}^Y$ with $I_x(y) = I(x, y)$. Forgetting now for a while about the data table, any fuzzy set $M \in \mathbf{L}^Y$ can be seen as a fuzzy set of attributes of some object with $M(y)$ being a degree to which the object has attribute y. For fuzzy set $M \in \mathbf{L}^Y$ of attributes, we define a *degree* $||A \Rightarrow B||_M \in L$ *to which* $A \Rightarrow B$ *is valid in* M by

$$||A \Rightarrow B||_M = S(A, M)^* \rightarrow S(B, M). \quad (13)$$

It is easily seen that if M is a fuzzy set of attributes of some object x then $||A \Rightarrow B||_M$ is the degree to which "if it is (very) true that x has all attributes from A then x has all attributes from B". For a system $\mathcal{M}$ of $\mathbf{L}$-sets in Y, define a degree $||A \Rightarrow B||_{\mathcal{M}}$ to which $A \Rightarrow B$ is true in (each M from) $\mathcal{M}$ by

$$||A \Rightarrow B||_{\mathcal{M}} = \bigwedge_{M\in\mathcal{M}} ||A \Rightarrow B||_M. \quad (14)$$

Finally, given a data table $\mathcal{T} = \langle X, Y, I\rangle$ and putting $\mathcal{M} = \{I_x \,|\, x \in X\}$, $||A \Rightarrow B||_{\mathcal{M}}$ is a degree to which it is true that $A \Rightarrow B$ is true in each row of table $\mathcal{T}$, i.e. a degree to which "for each object $x \in X$: if it is (very) true that x has all attributes from A, then x has all attributes from B". This degree is denoted by $||A \Rightarrow B||_{\langle X,Y,I\rangle}$ and is called a degree to which $A \Rightarrow B$ is true in data table $\langle X, Y, I\rangle$.

Therefore, we have two possible interpretations of our formulas $A \Rightarrow B$. First, given a data table $\mathcal{D}$ over domains with similarity relations, we can consider a truth degree $||A \Rightarrow B||_{\mathcal{D}}$ to which $A \Rightarrow B$ is true in $\mathcal{D}$. Second, given a data table $\mathcal{T}$ with fuzzy attributes, we can consider a truth degree $||A \Rightarrow B||_{\mathcal{T}}$ to which $A \Rightarrow B$ is true in $\mathcal{T}$.

Remark 4. A data table $\mathcal{T}$ with attributes can be considered as a data table $\mathcal{D}$ over domains with similarities. Namely, for each $y \in Y$, one can take $D_y = L$

(attribute values are truth degrees) and for $a, b \in L$ one can put $a \approx_y b = a \leftrightarrow b$ (this way, $\approx_y$ becomes an **L**-equality in D_y). Then, however, as one can easily see by a counterexample, we do not have $||A \Rightarrow B||_{\mathcal{T}} = ||A \Rightarrow B||_{\mathcal{D}}$ in general.

An interesting question is whether for a given data table $\mathcal{T}$ with fuzzy attributes, one can find a data table $\mathcal{D}_{\mathcal{T}}$ over domains with similarities such that $||A \Rightarrow B||_{\mathcal{T}} = ||A \Rightarrow B||_{\mathcal{D}_{\mathcal{T}}}$. And conversely, whether for a given data table $\mathcal{D}$ over domains with similarities, one can find a data table $\mathcal{T}_{\mathcal{D}}$ with fuzzy attributes such that $||A \Rightarrow B||_{\mathcal{D}} = ||A \Rightarrow B||_{\mathcal{T}_{\mathcal{D}}}$. That is, we are looking for table-transformation mappings $\mathcal{T} \mapsto \mathcal{D}_{\mathcal{T}}$ and $\mathcal{D} \mapsto \mathcal{T}_{\mathcal{D}}$ such that $\mathcal{T}$ and $\mathcal{D}_{\mathcal{T}}$ have the same valid formulas, and $\mathcal{D}$ and $\mathcal{T}_{\mathcal{D}}$ have the same valid formulas. In what follows we give a positive answer.

Suppose we are given $\mathcal{T} = \langle X, Y, I \rangle$. Define a data table $\mathcal{D}_{\mathcal{T}} = \langle X_{\mathcal{T}}, Y, \{\langle D_y, \approx_y \rangle \,|\, y \in Y\}, T \rangle$ over domains with similarities as follows:

- $X_{\mathcal{T}} = X \cup X'$ where $X' = \{x' \,|\, x \in X\}$ (i.e. $X \cap X' = \emptyset$ and $|X| = |X'|$),
- $D_y = X_{\mathcal{T}}$ for each $y \in Y$,
- for $x_1, x_2 \in D_y$, put

$$x_1 \approx_y x_2 = \begin{cases} 1 & \text{for } x_1 = x_2, \\ I(z_1, y) \wedge I(z_2, y) & \text{for } x_1 \neq x_2, x_i = z_i^{(\prime)} \text{ for } z_i \in X \ (i = 1, 2), \end{cases}$$

- $T(x, y) = x$ for $x \in X_{\mathcal{T}}$, $y \in Y$.

Here, $x = z^{(\prime)}$ means that either x is z, or x is z'. Then $\mathcal{D}_{\mathcal{T}}$ is a data table over domains with similarities.

Suppose we are given $\mathcal{D}_{\mathcal{T}} = \langle X, Y, \{\langle D_y, \approx_y \rangle \,|\, y \in Y\}, T \rangle$. Define a data table $\mathcal{T} = \langle X_{\mathcal{D}}, Y, I \rangle$ with fuzzy attributes as follows:

- $X_{\mathcal{D}} = X \times X$,
- $I(\langle x_1, x_2 \rangle, y) = (T(x_1, y) \approx_y T(x_2, y))$ for $\langle x_1, x_2 \rangle \in X_{\mathcal{D}}$, $y \in Y$.

Then $\mathcal{T}_{\mathcal{D}}$ is a data table over domains with similarities.

Theorem 1. *Let $\mathcal{T}$ be a data table with fuzzy attributes and $A, B \in \mathbf{L}^Y$. Then*

$$||A \Rightarrow B||_{\mathcal{T}} = ||A \Rightarrow B||_{\mathcal{D}_{\mathcal{T}}}. \tag{15}$$

Proof. The proof is technically involved, so we present just a sketch. First, it is easy to verify the following claims:

- $(a_1^* \to b_1) \wedge (a_2^* \to b_2) \leq (a_1 \wedge a_2)^* \to (b_1 \wedge b_2)$,
- $x_1(C) \approx x_2(C) = S(C, I_{z_1}) \wedge S(C, I_{z_2})$ for any $C \in \mathbf{L}^Y$ and any $x_1 \neq x_2$ such that $x_1 = z_1^{(\prime)}, x_2 = z_2^{(\prime)}$ for some $z_1, z_2 \in X$,
- $x(C) \approx x(C) = 1$ for $x \in X_{\mathcal{T}}$.

Now, we have

$$\begin{aligned} &||A \Rightarrow B||_{\mathcal{D}_{\mathcal{T}}} = \\ &= \textstyle\bigwedge_{x_1, x_2 \in X_{\mathcal{D}}} (x_1(A) \approx x_2(A))^* \to (x_1(B) \approx x_2(B)) = \mathfrak{A} \wedge \mathfrak{B} \wedge \mathfrak{C} \end{aligned}$$

where

$$\mathfrak{A} = \bigwedge_{x_1,x_2 \in X_{\mathcal{D}}, x_1 = x_2} (x_1(A) \approx x_2(A))^* \to (x_1(B) \approx x_2(B)) = 1,$$

$$\mathfrak{B} = \bigwedge_{x_1,x_2 \in X_{\mathcal{D}}, \{x_1,x_2\} = \{z,z'\}, z \in X} (x_1(A) \approx x_2(A))^* \to (x_1(B) \approx x_2(B)) =$$

$$= \bigwedge_{z \in X} (S(A, I_z) \wedge S(A, I_z))^* \to (S(B, I_z) \wedge S(B, I_z)) =$$

$$= \bigwedge_{z \in X} S(A, I_z)^* \to S(B, I_z) = ||A \Rightarrow B||_{\mathcal{T}},$$

and

$$\mathfrak{C} = \bigwedge_{\{x_1,x_2\} = \{z_1^{(\prime)}, z_2^{(\prime)}\}, z_1 \neq z_2 \in X} (x_1(A) \approx x_2(A))^* \to (x_1(B) \approx x_2(B)) =$$

$$= \bigwedge_{\{x_1,x_2\} = \{z_1^{(\prime)}, z_2^{(\prime)}\}, z_1 \neq z_2 \in X} (S(A, I_{z_1}) \wedge S(A, I_{z_2}))^* \to (S(B, I_{z_1}) \wedge S(B, I_{z_2})) \geq$$

$$= \bigwedge_{\{x_1,x_2\} = \{z_1^{(\prime)}, z_2^{(\prime)}\}, z_1 \neq z_2 \in X} [(S(A, I_{z_1})^* \to S(B, I_{z_1})) \wedge (S(A, I_{z_1})^* \to S(B, I_{z_1}))] =$$

$$= \bigwedge_{z \in X} S(A, I_z)^* \to S(B, I_z) = \mathfrak{B}.$$

Therefore,

$$||A \Rightarrow B||_{\mathcal{D}_{\mathcal{T}}} = \mathfrak{B} = ||A \Rightarrow B||_{\mathcal{T}},$$

completing the proof.

Remark 5. (1) Note that in the definition of $\mathcal{D}_{\mathcal{T}}$, $\approx_y$ need not be **L**-equality relations. However, $\approx_y$ is always an **L**-equivalence relation in D_y. In fact, it satisfies even $(x_1 \approx_y x_2) \wedge (x_2 \approx_y x_3) \leq (x_1 \approx_y x_3)$ (this immediately follows from the definition).

(2) However, if we would like $\mathcal{D}_{\mathcal{T}}$ with $\approx_y$ being **L**-equalities, we could take factor sets $D'_y = D_y/{\approx_y}$ of D_y by **L**-equivalences $\approx_y$ and take, instead of $\approx_y$ their factor relations $\approx'_y$ which are **L**-equalities in D'_y (details about these definitions can be found in [4]). For the thus obtained $\mathcal{D}'_{\mathcal{T}}$ we would have $||A \Rightarrow B||_{\mathcal{D}_{\mathcal{T}}} = ||A \Rightarrow B||_{\mathcal{D}'_{\mathcal{T}}}$ for any $A \Rightarrow B$.

Theorem 2. *Let $\mathcal{D}$ be a data table over domains with similarities and $A, B \in \mathbf{L}^Y$. Then*

$$||A \Rightarrow B||_{\mathcal{D}} = ||A \Rightarrow B||_{\mathcal{T}_{\mathcal{D}}}. \tag{16}$$

Proof. Using definitions and basic properties of residuated lattices with hedges, we have

$$||A \Rightarrow B||_{\mathcal{D}} = \bigwedge_{x_1,x_2 \in X} [x_1(A) \approx x_2(A)]^* \to [x_1(B) \approx x_2(B)] =$$

$$= \bigwedge_{x_1,x_2\in X}[\bigwedge_{y\in Y}(A(y) \to (T(x_1,y) \approx_y T(x_2,y)))]^* \to \to [\bigwedge_{y\in Y}(B(y) \to (T(x_1,y) \approx_y T(x_2,y)))] =$$
$$= \bigwedge_{x_1,x_2\in X}[\bigwedge_{y\in Y}(A(y) \to I(\langle x_1,x_2\rangle,y))]^* \to [\bigwedge_{y\in Y}(B(y) \to I(\langle x_1,x_2\rangle,y))] =$$
$$= \bigwedge_{x_1,x_2\in X}[\bigwedge_{y\in Y}(A(y) \to I_{\langle x_1,x_2\rangle}(y))]^* \to [\bigwedge_{y\in Y}(B(y) \to I_{\langle x_1,x_2\rangle}(y))] =$$
$$= \bigwedge_{x_1,x_2\in X}[S(A, I_{\langle x_1,x_2\rangle}]^* \to [S(B, I_{\langle x_1,x_2\rangle}] =$$
$$= ||A \Rightarrow B||_{\mathcal{T}_\mathcal{D}}.$$

3.3 Semantic entailment

Having now two semantics of functional dependencies $A \Rightarrow B$, i.e. two ways of interpreting $A \Rightarrow B$, we introduce the concepts of a model and semantic entailment of both of the semantics. The main result of this section says that although the two semantics are different, their concepts of semantic entailment coincide.

Consider an $\mathbf{L}$-set T of fuzzy functional dependencies. We will sometimes call T an $\mathbf{L}$-set T of fuzzy attribute implications, depending on the semantics we consider. From the point of view of logic, T can be seen as a theory, i.e. a degree $T(A \Rightarrow B)$ to which $A \Rightarrow B$ belongs to T can be seen as a degree to which we assume the validity $A \Rightarrow B$. This corresponds to the notion of a theory as a fuzzy set of axioms in fuzzy logic [19, 29]. From the user's point of view T can be seen a fuzzy set of functional dependencies on data such that $T(A \Rightarrow B)$ is a degree to which $A \Rightarrow B$ holds true in data. If T is crisp (which is particularly interesting) we write $A \Rightarrow B \in T$ if $T(A \Rightarrow B) = 1$ and $A \Rightarrow B \notin T$ if $T(A \Rightarrow B) = 0$.

For a fuzzy set T of fuzzy functional dependencies, the set $\mathrm{Mod}^{\mathrm{FD}}(T)$ of all *models* (or FD-models, for "models of T as functional dependencies") of T is defined by

$$\mathrm{Mod}^{\mathrm{FD}}(T) = \{\mathcal{D} \,|\, \text{for each } A, B \in \mathbf{L}^Y : T(A \Rightarrow B) \leq ||A \Rightarrow B||_\mathcal{D}\},$$

where $\mathcal{D}$ stands for an arbitrary data table over domains with similarities. That is, $\mathcal{D} \in \mathrm{Mod}^{\mathrm{FD}}(T)$ means that for each functional dependence $A \Rightarrow B$, a degree to which $A \Rightarrow B$ holds in $\mathcal{D}$ is higher than or at least equal to a degree $T(A \Rightarrow B)$ prescribed by T. Particularly, for a crisp T, $\mathrm{Mod}^{\mathrm{FD}}(T) = \{\mathcal{D} \,|\, \text{for each } A \Rightarrow B \in T : ||A \Rightarrow B||_\mathcal{D} = 1\}$.

A degree $||A \Rightarrow B||^{\mathrm{FD}}_T \in L$ to which $A \Rightarrow B$ *semantically follows* from a fuzzy set T of functional dependencies is defined by

$$||A \Rightarrow B||^{\mathrm{FD}}_T = \bigwedge_{\mathcal{D}\in\mathrm{Mod}^{\mathrm{FD}}(T)} ||A \Rightarrow B||_\mathcal{D}.$$

The same concepts can be introduced for the second semantics. For a fuzzy set T of fuzzy attribute implications, the set $\mathrm{Mod}^{\mathrm{AI}}(T)$ of all *models* (or AI-models, for "models of T as attribute implications") of T is defined by

$$\mathrm{Mod}^{\mathrm{AI}}(T) = \{M \in \mathbf{L}^Y \,|\, \text{for each } A, B \in \mathbf{L}^Y : T(A \Rightarrow B) \leq ||A \Rightarrow B||_M\}.$$

That is, $M \in \mathrm{Mod}^{\mathrm{AI}}(T)$ means that for each attribute implication $A \Rightarrow B$, a degree to which $A \Rightarrow B$ holds in M is higher than or at least equal to a degree $T(A \Rightarrow B)$ prescribed by T. Particularly, for a crisp T, $\mathrm{Mod}^{\mathrm{AI}}(T) = \{M \in \mathbf{L}^Y \,|\, \text{for each } A \Rightarrow B \in T : ||A \Rightarrow B||_M = 1\}$.

A degree $||A \Rightarrow B||^{\mathrm{AI}}(T)_T \in L$ to which $A \Rightarrow B$ *semantically follows* from a fuzzy set T of attribute implications is defined by

$$||A \Rightarrow B||_T^{\mathrm{AI}} = \bigwedge_{M \in \mathrm{Mod}^{\mathrm{AI}}(T)} ||A \Rightarrow B||_M.$$

Note that a model M of T can be seen as one-row table $\mathcal{T}_M$ with just one object (row) x with $I(x, y) = M(y)$.

Lemma 1. *For $A, B \in \mathbf{L}^Y$, a data table $\mathcal{D}$ over domains with similarities, and $c \in L$ we have*

$$c \leq ||A \Rightarrow B||_{\mathcal{D}} \text{ iff } ||A \Rightarrow c \otimes B||_{\mathcal{D}} = 1. \tag{17}$$

Proof. We have

$$\begin{aligned}
&c \to ||A \Rightarrow B||_{\mathcal{D}} = \\
&= c \to \textstyle\bigwedge_{x_1,x_2 \in X}((x_1(A) \approx x_2(A))^* \to (x_1(B) \approx x_2(B))) = \\
&= \textstyle\bigwedge_{x_1,x_2 \in X}(c \to ((x_1(A) \approx x_2(A))^* \to (x_1(B) \approx x_2(B)))) = \\
&= \textstyle\bigwedge_{x_1,x_2 \in X}((x_1(A) \approx x_2(A))^* \to (c \to (x_1(B) \approx x_2(B)))) = \\
&= \textstyle\bigwedge_{x_1,x_2 \in X}((x_1(A) \approx x_2(A))^* \to (c \to (\bigwedge_{y \in Y}[B(y) \to (T(x_1,y) \approx_y T(x_2,y))]))) \\
&= \textstyle\bigwedge_{x_1,x_2 \in X}((x_1(A) \approx x_2(A))^* \to (\bigwedge_{y \in Y} c \to [B(y) \to (T(x_1,y) \approx_y T(x_2,y))]) \\
&= \textstyle\bigwedge_{x_1,x_2 \in X}((x_1(A) \approx x_2(A))^* \to (\bigwedge_{y \in Y}[c \otimes B(y) \to (T(x_1,y) \approx_y T(x_2,y))]) \\
&= ||A \Rightarrow c \otimes B||_{\mathcal{D}}.
\end{aligned}$$

Therefore, we have $c \leq ||A \Rightarrow B||_{\mathcal{D}}$ iff $c \to ||A \Rightarrow B||_{\mathcal{D}} = 1$ iff $||A \Rightarrow c \otimes B||_{\mathcal{D}} = 1$.

Lemma 1 has important consequences. It enables us to reduce the concept of a model of a fuzzy set of functional dependencies (attribute implications) to the concept of a model of an ordinary set of functional dependencies (attribute implications), and to reduce the concept of semantic entailment from a fuzzy set of functional dependencies (attribute implications) to the concept of semantic entailment from an ordinary set of functional dependencies (attribute implications):

Lemma 2. *Let T be a fuzzy set of functional dependencies and $A, B \in \mathbf{L}^Y$. Define an ordinary set $\mathrm{c}(T)$ of functional dependencies by*

$$\mathrm{c}(T) = \{A \Rightarrow T(A \Rightarrow B) \otimes B \,|\, A, B \in \mathbf{L}^Y \text{ and } T(A \Rightarrow B) \otimes B \neq \emptyset\}. \tag{18}$$

Then we have

$$\mathrm{Mod}^{\mathrm{FD}}(T) = \mathrm{Mod}^{\mathrm{FD}}(\mathrm{c}(T)), \tag{19}$$

$$||A \Rightarrow B||_T^{\mathrm{FD}} = ||A \Rightarrow B||_{\mathrm{c}(T)}^{\mathrm{FD}}. \tag{20}$$

Proof. (19) directly using Lemma 1. (20) is a consequence of (19).

A similar lemma concerning the other semantics was proved in [13].

Furthermore, Lemma 1 enables us to reduce the concept of a degree of entailment of a functional dependence from a fuzzy set of functional dependencies to the concept of an entailment in degree 1 (full entailment) of a functional dependence from a fuzzy set of functional dependencies:

Lemma 3. *For $A, B \in \mathbf{L}^Y$ and a fuzzy set T of fuzzy attribute implications we have*

$$||A \Rightarrow B||_T^{\mathrm{FD}} = \bigvee\{c \in L \mid ||A \Rightarrow c \otimes B||_T^{\mathrm{FD}} = 1\}.$$

Proof. Easy using Lemma 1.

Therefore, we have:

Corollary 1. *For $A, B \in \mathbf{L}^Y$ and a fuzzy set T of functional dependencies we have*

$$||A \Rightarrow B||_T^{\mathrm{FD}} = \bigvee\{c \in L \mid ||A \Rightarrow c \otimes B||_{\mathrm{c}(T)}^{\mathrm{FD}} = 1\},$$

with $\mathrm{c}(T)$ *defined by* (18).

Corollary 1 shows that the concept of a degree of entailment from a fuzzy set of functional dependencies can be reduced to entailment in degree 1 from a set of functional dependencies.

The following theorem is the main result of this section.

Theorem 3. *For any fuzzy set T of functional dependencies and any functional dependence we have*

$$||A \Rightarrow B||_T^{\mathrm{FD}} = ||A \Rightarrow B||_T^{\mathrm{AI}}. \tag{21}$$

Proof. The proof goes by proving both $||A \Rightarrow B||_T^{\mathrm{FD}} \leq ||A \Rightarrow B||_T^{\mathrm{AI}}$ and $||A \Rightarrow B||_T^{\mathrm{FD}} \geq ||A \Rightarrow B||_T^{\mathrm{AI}}$. To see the first inequality, it is enough to show that for each $M \in \mathrm{Mod}^{\mathrm{AI}}(T)$ there is $\mathcal{D} \in \mathrm{Mod}^{\mathrm{FD}}(T)$ such that $||A \Rightarrow B||_M = ||A \Rightarrow B||_{\mathcal{D}}$. For this purpose, consider $\mathcal{D} = \mathcal{D}_{\mathcal{T}_M}$ where $\mathcal{T}_M$ is the one-row data table corresponding to M (see above). Then, Theorem 1 implies that since M is a model of T, $\mathcal{D}$ is a model of T as well, and furthermore, that $||A \Rightarrow B||_M = ||A \Rightarrow B||_{\mathcal{D}}$. A proof of the second inequality is similar.

3.4 Complete systems of deduction rules for functional dependencies

In this section, we introduce an axiomatic system for reasoning with functional dependencies and prove completeness theorems. Our basic technique is a reduction (based on the results from previous sections) to fuzzy attribute logic [12, 13] for which there is a complete system of deduction rules.

Our axiomatic system consists of the following *deduction rules*.

(Ax) infer $A \cup B \Rightarrow A$,
(Cut) from $A \Rightarrow B$ and $B \cup C \Rightarrow D$ infer $A \cup C \Rightarrow D$,
(Mul) from $A \Rightarrow B$ infer $c^* \otimes A \Rightarrow c^* \otimes B$

for each $A, B, C, D \in \mathbf{L}^Y$, and $c \in L$. Rules (Ax)–(Mul) are to be understood as follows: having functional dependencies which are of the form of functional dependencies in the input part (the part preceding "infer") of a rule, a rule allows us to infer (in one step) the corresponding functional dependence in the output part (the part following "infer") of a rule.

A fuzzy attribute implication $A \Rightarrow B$ is called *provable* from a set T of fuzzy attribute implications using a set $\mathcal{R}$ of deduction rules, written $T \vdash_{\mathcal{R}} A \Rightarrow B$, if there is a sequence $\varphi_1, \ldots, \varphi_n$ of fuzzy attribute implications such that φ_n is $A \Rightarrow B$ and for each φ_i we either have $\varphi_i \in T$ or φ_i is inferred (in one step) from some of the preceding formulas (i.e., $\varphi_1, \ldots, \varphi_{i-1}$) using some deduction rule from $\mathcal{R}$. If $\mathcal{R}$ consists of (Ax)–(Mul), we say just "provable ..." instead of "provable ... using $\mathcal{R}$" and write just $T \vdash A \Rightarrow B$ instead of $T \vdash_{\mathcal{R}} A \Rightarrow B$.

Theorem 4 (completeness). *Let* $\mathbf{L}$ *and* Y *be finite. Let* T *be a set of functional dependencies. Then*

$$T \vdash A \Rightarrow B \quad \textit{iff} \quad ||A \Rightarrow B||_T^{\mathrm{FD}} = 1.$$

Proof. By Theorem 3, $||A \Rightarrow B||_T^{\mathrm{FD}} = 1$ iff $||A \Rightarrow B||_T^{\mathrm{AI}} = 1$. Furthermore, as shown in [13], $||A \Rightarrow B||_T^{\mathrm{AI}} = 1$ iff $T \vdash A \Rightarrow B$, hence the result.

Now, we are going to define a notion of a degree $|A \Rightarrow B|_T$ of provability of a functional dependence of a fuzzy set T of functional dependencies. Then, we show that $|A \Rightarrow B|_T = ||A \Rightarrow B||_T^{\mathrm{FD}}$ which can be understood as a graded completeness (completeness in degrees). Note that graded completeness was introduced by Pavelka [29], see also [19, 23] for further information.

For a fuzzy set T of functional dependencies and for $A \Rightarrow B$ we define a *degree* $|A \Rightarrow B|_T \in L$ *to which* $A \Rightarrow B$ *is provable from* T by

$$|A \Rightarrow B|_T = \bigvee \{c \in L \mid \mathrm{c}(T) \vdash A \Rightarrow c \otimes B\}, \tag{22}$$

where $\mathrm{c}(T)$ is defined by (18).

Theorem 5 (graded completeness). *Let* $\mathbf{L}$ *and* Y *be finite. Then for every fuzzy set* T *of functional dependencies and* $A \Rightarrow B$ *we have* $|A \Rightarrow B|_T = ||A \Rightarrow B||_T$.

Proof. Follows from Corollary 1 and Theorem 4.

In [13], we showed other systems of deduction rules which are equivalent to (Ax)–(Mul). Just for illustration, one of them consists of the following rules:

(Ax') infer $A \Rightarrow S(B, A) \otimes B$,
(Wea') from $A \Rightarrow B$ infer $A \cup C \Rightarrow B$,
(Cut') from $A \Rightarrow e \otimes B$ and $B \cup C \Rightarrow D$ infer $A \cup C \Rightarrow e^* \otimes D$

for each $A, B, C, D \in \mathbf{L}^Y$, and $e \in L$.

Direct proof of completeness

Since the results of [12, 13] to which we refer in the proof of Theorem 4 (completeness w.r.t. semantics given by data tables with fuzzy attributes) were not published yet, we now give a sketch of a direct proof of Theorem 4.

A set T of functional dependencies is said to be *syntactically closed* if $T \vdash A \Rightarrow B$ iff $A \Rightarrow B \in T$, i.e. if $T = \{A \Rightarrow B \mid T \vdash A \Rightarrow B\}$. A set T of functional dependencies is said to be *semantically closed* if $||A \Rightarrow B||_T^{\mathrm{FD}} = 1$ iff $A \Rightarrow B \in T$, i.e. if $T = \{A \Rightarrow B \mid ||A \Rightarrow B||_T^{\mathrm{FD}} = 1\}$.

By standard logical arguments one can see that Theorem 4 is a consequence of the following assertion: A set T of functional dependencies is syntactically closed iff T is semantically closed. The fact that if T is semantically closed then T is syntactically closed can be proved by standard arguments observing that each the rules (Ax)–(Mul) is sound. Thus to prove completeness, it remains to see the following lemma.

Lemma 4. *Let T be a set of functional dependencies, let both Y and L be finite. If T is syntactically closed then T is semantically closed.*

Proof. Sketch: Let T be syntactically closed. In order to show that T is semantically closed, it suffices to show $\{A \Rightarrow B \mid ||A \Rightarrow B||_T^{\mathrm{FD}} = 1\} \subseteq T$. We prove this by showing that if $A \Rightarrow B \notin T$ then $A \Rightarrow B \notin \{A \Rightarrow B \mid ||A \Rightarrow B||_T^{\mathrm{FD}} = 1\}$. It can be checked that the usual rules which result from Armstrong axioms (and thus also their consequences) by replacing sets with fuzzy sets can be derived from (Ax)–(Mul). Hence, since T is syntactically closed, it is closed under all of these rules.

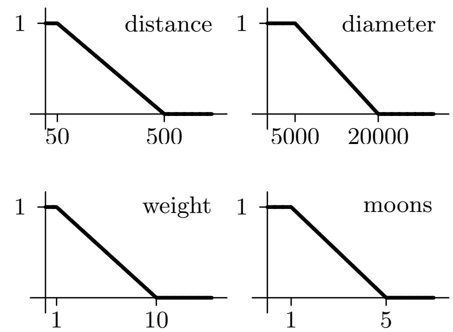
Let thus $A \Rightarrow B \notin T$. To see $A \Rightarrow B \notin \{A \Rightarrow B \mid ||A \Rightarrow B||_T^{\mathrm{FD}} = 1\}$, we show that there is $\mathcal{D} \in \mathrm{Mod}^{\mathrm{FD}}(T)$ which is not a model of $A \Rightarrow B$. For this purpose, one can take $\mathcal{D} = \mathcal{D}_{\mathcal{T}_{A^+}}$ where $A^+ \in \mathbf{L}^Y$ is the largest one such that $A \Rightarrow A^+ \in T$. We omit the rest of the proof and leave it to a full version of this paper.

4 Examples

For illustration, consider data table $\mathcal{D} = \langle X, Y, T \rangle$ from Tab. 1 (left) which describes basic properties of planets of our solar system. The table contains the following attributes: *name of the planet*, *distance from sun* (in thousands of kilometers), *equatorial diameter* (in kilometers), *weight* (in weights of Earth), *number of known moons*. By a quick inspection of the table one can see that there are no "interesting" functional dependencies (in the classical sense). On the other hand, it is intuitively acceptable to say, e.g., "if two planets have more or less the same number of moons, then they will have (almost) the same weight". We now demonstrate that if we equip the data table with a collection of suitable similarity relations $\{\langle D_y, \approx_y \rangle \mid y \in Y\}$, the implication between number of moons and weight is expressed by a fuzzy attribute implication which is true in $\mathcal{D} = \langle X, Y, \{\langle D_y, \approx_y \rangle \mid y \in Y\}, T \rangle$.

Table 1. Data table $\mathcal{D} = \langle X, Y, \{\langle D_y, \approx_y\rangle \mid y \in Y\}, T\rangle$

name	dist.	diam.	weight	moons
Mercury	57.9	4878	0.056	0
Venus	108.2	12103	0.815	0
Earth	149.6	12714	1.000	1
Mars	227.9	6787	0.107	2
Jupiter	778.3	134700	317.700	39
Saturn	1427.0	120000	95.200	30
Uranus	2870.0	50800	14.660	21
Neptune	4496.7	48600	17.230	8
Pluto	5900.0	2300	0.002	1



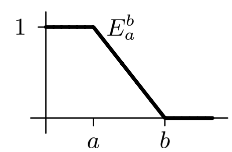


Fig. 1. E_a^b

First, we define a similarity relation $\approx_y$ on the domain D_y of each attribute $y \in Y$. A common way to define similarity is the following: we consider two objects similar to a degree to which is it true that the objects are "close" (e.g., close in terms of their distance). So, let $\mathbf{L}$ be a residuated lattice on real unit interval with globalization and denote by E_a^b an $\mathbf{L}$-set in $[0, \infty)$ defined by

$$E_a^b(x) = \begin{cases} 1 & \text{if } x < a, \\ \frac{b-x}{b-a} & \text{if } x \geq a \text{ and } x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

The shape of E_a^b is depicted in Fig. 1. E_a^b expresses that if the distance between two reals drops below a, then the reals are indistinguishable (with respect to E_a^b); if the distance exceeds b, the reals are fully distinct (with respect to E_a^b); reals with distances between a and b are given proportional truth degrees between 1 and 0. Thus, for any real numbers x_1 and x_2 we can define their E_a^b-similarity degree to be $E_a^b(|x_1 - x_2|)$, i.e. the degree to which $|x_1 - x_2|$ belongs to E_a^b. Put

$$\begin{aligned} x_1 \approx_s x_2 &= E_{50}^{500}(|x_1 - x_2|), \\ x_1 \approx_d x_2 &= E_{5000}^{20000}(|x_1 - x_2|), \\ x_1 \approx_w x_2 &= E_{1}^{10}(|x_1 - x_2|), \\ x_1 \approx_m x_2 &= E_{1}^{5}(|x_1 - x_2|), \end{aligned}$$

where $s \in Y$ denotes distance from sun, $d \in Y$ denotes diameter, $w \in Y$ denotes weight, and $m \in Y$ denotes number of moons, see Tab. 1 (right) for the corresponding shapes. Now we can consider validity of functional dependencies in data table $\mathcal{D} = \langle X, Y, \{\langle D_y, \approx_y\rangle \mid y \in Y\}, T\rangle$. For instance, $\{{}^{0.7}\!/m\} \Rightarrow \{{}^{0.6}\!/d, w\}$, saying "if the numbers of moons are similar in degree (at least) 0.7, then the diameters and similar in degree 0.6 and weights are similar in degree 1", is true in $\mathcal{D}$. If we take into account the meaning of $\approx_y$'s, the formula can be read:

"if $|x[m] - x[m]| \leq 2$ then $|x[d] - x'[d]| \leq 11000$ and $|x[w] - x'[w]| \leq 1$",

i.e., the implication says: "if the difference between numbers of moons of x and x' is less than or equal to 2 then the difference between their diameters is at most 11000 and the difference between their weights in at most one weight of Earth).

Recall that validity of fuzzy functional dependencies in $\mathcal{D}$ can be expressed by validity of fuzzy attribute implications in the corresponding data table $\mathcal{T}_{\mathcal{D}}$ with fuzzy attributes induced by $\mathcal{D}$, see Theorem 2. Table $\mathcal{T}_{\mathcal{D}}$ for a data table from Tab. 1 can is shown in Tab. 2. Note that for technical reasons (namely, the algorithm mentioned below requires a finite scale L of truth degrees), we rounded the exact values of $L = [0, 1]$ from $\mathcal{T}_{\mathcal{D}}$ down to values of $L = \{0, 0.1, 0.2, \ldots, 0.9, 1\}$. Furthermore, we removed duplicit and trivial rows in Tab. 2.

Using the results of [7, 11], we can compute a minimal basis of all fuzzy attribute implications from $\mathcal{T}_{\mathcal{D}}$. Recall that a minimal basis is a set T of attribute implications true (in degree 1) in $\mathcal{T}_{\mathcal{D}}$ such that (1) each attribute implication which is true in $\mathcal{T}_{\mathcal{D}}$ semantically follows from T, and (2) no attribute implication can be removed from T without violating (1). Due to Theorem 3 a minimal basis of $\mathcal{T}_{\mathcal{D}}$ is also a minimal basis of $\mathcal{D}$. It is clear that what we just described, is true for any $\mathcal{D}$: Theorem 2, Theorem 3, and the algorithm for computation of a minimal basis of attribute implications from data tables with fuzzy attributes described in [7] give an algorithm for computation of a minimal basis of $\mathcal{D}$ (the algorithm has a polynomial time delay complexity). In our case, the basis contains the following formulas:

$$\begin{array}{ll}
\{s, {}^{0.8\!}/d, w, m\} \Rightarrow \{d\}, & \{{}^{0.9\!}/s, {}^{0.8\!}/d, w, {}^{0.7\!}/m\} \Rightarrow \{m\}, \\
\{{}^{0.8\!}/s, d, w, {}^{0.7\!}/m\} \Rightarrow \{s, m\}, & \{{}^{0.7\!}/s, {}^{0.8\!}/d, w, m\} \Rightarrow \{{}^{0.9\!}/s\}, \\
\{{}^{0.1\!}/s\} \Rightarrow \{{}^{0.7\!}/s, {}^{0.8\!}/d, w, {}^{0.7\!}/m\}, & \{{}^{0.7\!}/d, {}^{0.8\!}/w\} \Rightarrow \{{}^{0.8\!}/d\}, \\
\{{}^{0.6\!}/d, w, {}^{0.8\!}/m\} \Rightarrow \{m\}, & \{{}^{0.6\!}/d, {}^{0.9\!}/w\} \Rightarrow \{w, {}^{0.7\!}/m\}, \\
\{{}^{0.4\!}/d\} \Rightarrow \{{}^{0.6\!}/d, {}^{0.8\!}/w\}, & \{{}^{0.1\!}/d\} \Rightarrow \{{}^{0.3\!}/d\}, \\
\{{}^{0.1\!}/w\} \Rightarrow \{{}^{0.6\!}/d, {}^{0.8\!}/w\}, & \{{}^{0.1\!}/m\} \Rightarrow \{{}^{0.6\!}/d, w, {}^{0.7\!}/m\}.
\end{array}$$

A fuzzy functional dependency $A \Rightarrow B$ holds in $\mathcal{D}$ in degree to which follows (syntactically/semantically) from the above-mentioned functional dependencies. One can see that all of the functional dependencies of the basis have a natural meaning in the data table $\mathcal{D}$.

5 Concluding remarks and future research

Our future research will focus on:

Table 2. Data table $\mathcal{T}_{\mathcal{D}}$ (after reduction)

	s	d	w	m
1.	0.0	0.0	0.0	0.0
2.	0.0	0.3	0.0	0.0
3.	0.0	0.6	1.0	1.0
4.	0.0	1.0	0.8	0.0
5.	0.0	1.0	1.0	1.0
6.	0.7	1.0	1.0	0.7
7.	0.8	0.9	1.0	0.7
8.	0.9	0.8	1.0	1.0
9.	0.9	0.9	1.0	1.0

- algorithms for various problems of fuzzy functional dependencies (for instance, a direct algorithm for the computation of a basis of fuzzy functional dependencies), [28] is a good survey of problems and algorithms in classical databases,
- further types of data dependencies in a fuzzy setting, like multivalued dependencies, etc., see [28],
- theoretical issues of relational databases with domains equipped with similarities (design, queries, etc.),
- implementation issues in relational databases with domains equipped with similarities,
- comparison with other approaches to database from the point of view of fuzzy logic [14, 27, 30, 32, 36],
- comparison with approaches to dependencies in other frameworks for uncertainty management, e.g. probabilistic dependencies [15],
- further approaches to data tables over domains with similarities. For instance, our tables $\mathcal{D}$ are ordinary relations: each row x represents an existing association between the corresponding values $x[y]$ $(y \in Y)$. It might be interesting to consider $\mathcal{D}$ as a fuzzy relation, i.e. to consider for each row x a degree $r(x)$ to which values $x[y]$ $(y \in Y)$ are associated. In fact, such an approach was proposed in [33] where one can find motivating examples. We have proved results analogous to the results of the present paper also for this more general case ($\mathcal{D}$ as a fuzzy relation). Due to a lack of space, we do not include the results here and postpone their presentation to a forthcoming paper.

References

1. Abiteboui S. *et al.*: The Lowell database research self-assessment. *Communications of ACM* **48**(5)(2005), 111–118.
2. Armstrong W. W.: Dependency structures in data base relationships. *IFIP Congress*, Geneva, Switzerland, 1974, pp. 580–583.
3. Bělohlávek R.: Similarity relations in concept lattices. *J. Logic Comput.* 10(6):823–845, 2000.

4. Bělohlávek R.: *Fuzzy Relational Systems: Foundations and Principles.* Kluwer, Academic/Plenum Publishers, New York, 2002.
5. Bělohlávek R.: Concept lattices and order in fuzzy logic. *Ann. Pure Appl. Logic* **128**(2004), 277–298.
6. Bělohlávek R.: Algorithms for fuzzy concept lattices. *Proc. Fourth Int. Conf. on Recent Advances in Soft Computing.* Nottingham, United Kingdom, 12-13 December, 2002, pp. 200–205.
7. Bělohlávek R., Chlupová M., Vychodil V.: Implications from data with fuzzy attributes. AISTA 2004 in Cooperation with the IEEE Computer Society Proceedings, 2004, 5 pages, ISBN 2–9599776–8–8.
8. Bělohlávek R., Funioková T., Vychodil V.: Fuzzy closure operators with truth stressers. *Logic Journal of IGPL* (to appear).
9. Bělohlávek R., Vychodil V.: Implications from data with fuzzy attributes vs. scaled binary attributes. In: FUZZ-IEEE 2005, The IEEE International Conference on Fuzzy Systems, May 22–25, 2005, Reno (Nevada, USA), pp. 1050–1055 (proceedings on CD), abstract in printed proceedings, p. 53, ISBN 0–7803–9158–6.
10. Bělohlávek R., Vychodil V.: Reducing the size of fuzzy concept lattices by hedges. In: FUZZ-IEEE 2005, The IEEE International Conference on Fuzzy Systems, May 22–25, 2005, Reno (Nevada, USA), pp. 663–668 (proceedings on CD), abstract in printed proceedings, p. 44, ISBN 0–7803–9158–6.
11. Bělohlávek R., Vychodil V.: Fuzzy attribute logic: attribute implications, their validity, entailment, and non-redundant basis. In: Yingming Liu, Guoqing Chen, Mingsheng Ying (Eds.): *Fuzzy Logic, Soft Computing & Computational Intelligence: Eleventh International Fuzzy Systems Association World Congress* (Vol. I), 2005, pp. 622–627. Tsinghua University Press and Springer, ISBN 7–302–11377–7.
12. Bělohlávek R., Vychodil V.: Fuzzy attribute logic: syntactic entailment and completeness. *Proceedings of the 8th Joint Conference on Information Sciences*, 2005, pp. 78–81, ISBN 0–9707890–3–3.
13. Bělohlávek R., Vychodil V.: Axiomatizations of fuzzy attribute logic. IICAI 2005 (Pune, India, to appear).
14. Buckles B. P., Petry F. E.: Fuzzy databases in the new era. Proceedings of the 1995 ACM symposium on Applied computing, pp. 497–502, Nashville, Tennessee, ISBN 0-89791-658-1, 1995.
15. Butz C. J., Wong S. K. M., Yao Y. Y.: On data and probabilistic dependencies. In IEEE Conf. Electrical and Computer Engineering, IEEE Press, Montreal, 1999, pp.1692–1697.
16. Cubero J. C., Vila M. A.: A new definition of fuzzy functional dependency in fuzzy relational datatabses. *Int. J. Intelligent Systems* **9**(5)(1994), 441–448.
17. Ganter B.: *Begriffe und Implikationen*, manuscript, 1998.
18. Ganter B., Wille R.: *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin, 1999.
19. Gerla G.: *Fuzzy Logic. Mathematical Tools for Approximate Reasoning.* Kluwer, Dordrecht, 2001.
20. Goguen J. A.: L-fuzzy sets. *J. Math. Anal. Appl.* **18**(1967), 145–174.
21. Goguen J. A.: The logic of inexact concepts. *Synthese* **18**(1968-9), 325–373.
22. Guigues J.-L., Duquenne V.: Familles minimales d'implications informatives resultant d'un tableau de données binaires. *Math. Sci. Humaines* **95**(1986), 5–18.
23. Hájek P.: *Metamathematics of Fuzzy Logic.* Kluwer, Dordrecht, 1998.
24. Hájek P.: On very true. *Fuzzy Sets and Systems* **124**(2001), 329–333.
25. Johnson D. S., Yannakakis M., Papadimitrou C. H.: On generating all maximal independent sets. *Inf. Processing Letters* **15**(1988), 129–133.

26. Klir G. J., Yuan B.: *Fuzzy Sets and Fuzzy Logic. Theory and Applications.* Prentice Hall, 1995.
27. Li D., Liu D.: *A Fuzzy Prolog Database System.* Taunton, England: Research Studies Press, 1990.
28. Maier D.: *The Theory of Relational Databases.* Computer Science Press, Rockville, 1983.
29. Pavelka J.: On fuzzy logic I, II, III. *Z. Math. Logik Grundlagen Math.* **25**(1979), 45–52, 119–134, 447–464.
30. Petry F.: *Fuzzy Databases: Principles and Applications.* Kluwer Academic, 1996.
31. Pollandt S.: *Fuzzy Begriffe.* Springer-Verlag, Berlin/Heidelberg, 1997.
32. Prade H., Testemale C.: Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries. *Information Sciences* **34**(1984), 115–143.
33. Raju K. V. S. V. N., Majumdar A. K.: Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems. *ACM Trans. Database Systems* Vol. 13, No. 2, 1988, pp. 129–166.
34. Takeuti G., Titani S.: Globalization of intuitionistic set theory. *Annals of Pure and Applied Logic* **33**(1987), 195–211.
35. Tyagi B. K., Sharfuddin A., Dutta R. N., Tayal D. K.: A complete axiomatization of fuzzy functional dependencies using fuzzy function. *Fuzzy Sets and Systems* **151**(2)(2005), 363–379.
36. Yang Q. *et al.*: Efficient processing of nested fuzzy SQL queries in a fuzzy database. *IEEE Trans. Knowledge and Data Eng.* Vol. 13, No. 6, 2001, pp. 884–901.
37. Zhang C., Zhang S.: *Association Rule Mining. Models and Algorithms.* Springer, Berlin, 2002.