# Data Tables with Similarity Relations: Functional Dependencies, Complete Rules and Non-redundant Bases*

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 rules $A \Rightarrow B$ describing attribute dependencies in tables over domains with similarity relations. $A \Rightarrow B$ reads "for any two table rows: similar values of attributes from $A$ imply similar values of attributes from $B$". The rules generalize ordinary functional dependencies in that they allow for processing of similarity of attribute values. Similarity is modeled by reflexive and symmetric fuzzy relations. We show a system of Armstrong-like derivation rules and prove its completeness (two versions). Furthermore, we describe a non-redundant basis of all rules which are true in a data table and present an algorithm to compute bases.

## 1 Introduction and Related Work

*Introduction.* Rules of the form $A \Rightarrow B$ where $A$ and $B$ are collections of attributes have been studied in several areas of computer science. We are interested in their role as describing dependencies known as functional dependencies [2, 12]. The interpretation of an ordinary functional dependence $A \Rightarrow B$ in a given data table $\mathcal{D}$ is the following: any two table rows in $\mathcal{D}$ which have the same values of attributes from $A$ have also the same values of attributes from $B$.

In a paper by 29 leading experts in database systems [1], it has been pointed out that one of the important future topics in database research is management of uncertainty. In particular, one should extend existing tools to allow for imprecision. For instance, not only exact matches but also approximate matches of data items, i.e. matches w.r.t. some underlying similarity, 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 take into account 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 $B$.

* 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.

M.L. Lee, K.L. Tan, and V. Wuwongse (Eds.): DASFAA 2006, LNCS 3882, pp. 644–658, 2006.

**Table 1.** Data table: there are no non-trivial ordinary functional dependencies but there are approximate dependencies

| | 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 |

As an illustrative example, consider a table from Tab. 1 describing planets of our solar system. The table contains the following attributes: *distance from sun* (in thousands of kilometers), *equatorial diameter* (in kilometers), *weight* (in weights of Earth), *number of known moons*; and objects *Mercury*, *Venus*, ... As one can see, there are no non-trivial ordinary functional dependencies in the data table. However, one can see that with an intuitive notion of similarity, there are dependencies saying that similar values of some attributes imply similar values in other attributes. For instance, similar distance from sun implies similar number of moons. On the other hand, Uranus and Neptune serve as a counterexample to a dependency saying that similar diameter implies similar number of moons. Needless to say, a precise meaning the above described dependencies depends on the definition of the similarity relations and the definition of validity of a functional dependency involving similarity relations. We come back to this example in Section 6.

One can think of many other examples of functional dependencies over domains with similarity relations and there is a question of an appropriate framework to put this idea into work. A feasible option is offered by fuzzy logic [11]. 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.

In the present paper, we introduce a concept of a functional dependence and its interpretation in data tables over domains with similarities. We present a system of axioms (deduction rules) and show its completeness as well as its graded

completeness. We describe non-reduntant bases of all functional dependencies which are true in a data table and present an algorithm for their computation.

*Related Work.* For an overview of modeling uncertainty and imprecision in data engineering and databases we refer to [6]. Various aspects of functional dependencies over domains equipped with similarity relations have already been studied, see e.g. [13, 14, 15], a good overview is [15]. Compared to our notion of a functional dependence and its validity, neither of these approaches does allow for using thresholds (see above). Therefore, our dependencies have more expressive capability. For instance, we can have dependencies like "age similarity in degree at least 0.7 and income similarity in degree at least 0.9 implies similarity in life insurance costs in degree 0.5" which is quite reasonable rule since it captures the possibly different influences of age and income on the conclusion concerning similarity in life insurance costs. Furthermore, we describe bases and an algorithm for their computation which the above-cited works did not.

## 2 Preliminaries

Fuzzy logic and fuzzy set theory are formal frameworks for a manipulation of a particular form of imperfection called fuzziness (vagueness). For an introduction to fuzzy logic we refer to [3, 9, 11]. In this section, we recall some concepts we need.

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 us to consider intermediate truth degrees of propositions, e.g. "$x_1$ is similar to $x_2$" has a truth degree 0.8, indicating that the proposition is almost true. In addition to $L$, one has to pick an appropriate collection of logical connectives (implication, conjunction, ...). A general choice covering almost all particular structures used in applications is a complete residuated lattice with a truth-stressing hedge (shortly, a hedge) [9], i.e. a structure $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, {}^*, 0, 1\rangle$ 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; $\otimes$ is commutative, associative, and $a \otimes 1 = 1 \otimes a = a$ for each $a \in L$; $\otimes$ and $\rightarrow$ satisfy so-called adjointness: $a \otimes b \leq c$ iff $a \leq b \rightarrow c$ for each $a, b, c \in L$; hedge ${}^*$ satisfies $1^* = 1$, $a^* \leq a$, $(a \rightarrow b)^* \leq a^* \rightarrow b^*$, $a^{**} = a^*$. 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". The above properties have natural logical interpretations [9].

A common choice of $\mathbf{L}$ is a structure with $L = [0, 1]$ (unit interval), $\wedge$ and $\vee$ being minimum and maximum. Three most important pairs of $\otimes$ and $\rightarrow$ are Łukasiewicz: $a \otimes b = \max(a + b - 1, 0)$, $a \rightarrow b = \min(1 - a + b, 1)$; Gödel (minimum): $a \otimes b = \min(a, b)$, $a \rightarrow b = 1$ for $a \leq b$ and $a \rightarrow b = b$ for $a > b$; Goguen (product): $a \otimes b = a \cdot b$, $a \rightarrow b = 1$ for $a \leq b$ and $a \rightarrow b = \frac{b}{a}$ for $a > b$. Another common choice is $L = \{a_0 = 0, a_1, \ldots, a_n = 1\} \subseteq [0, 1]$ $(a_0 < \cdots < a_n)$ 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: $a^* = 1$ for $a = 1$, and $a^* = 0$ for $a \neq 1$.

Having $\mathbf{L}$, we define usual notions: an $\mathbf{L}$-set (fuzzy set) $A$ in universe $U$ is a mapping $A\colon U \to L$, $A(u)$ 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 write $\{\ldots, u, \ldots\}$ instead of $\{\ldots, {}^{1}/u, \ldots\}$. Let $\mathbf{L}^U$ denote the collection of all $\mathbf{L}$-sets in $U$. Operations with $\mathbf{L}$-sets are defined in the usual way, i.e. componentwise [11]. An $\mathbf{L}$-set $A \in \mathbf{L}^X$ is called crisp if $A(x) \in \{0,1\}$ for each $x \in X$. Crisp $\mathbf{L}$-sets can be identified with ordinary sets. 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

$$S(A,B) = \bigwedge\nolimits_{u \in U} \big(A(u) \to B(u)\big), \tag{1}$$

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 $\mathbf{L}$-relation $\approx$ in $U$, i.e. a mapping $\approx\colon U \times U \to L$, 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 $(u \approx v) \otimes (v \approx w) \leq (u \approx w)$; $\mathbf{L}$-equivalence if it is reflexive, symmetric, and transitive. We will use reflexive and symmetric $\mathbf{L}$-relations to represent similarity on domains of attribute values.

Throughout the rest of the paper, $\mathbf{L}$ denotes an arbitrary complete residuated lattice with a hedge.

## 3 Functional Dependencies over Domains with Similarity Relations

### 3.1 Functional Dependencies and Their Validity

Suppose $Y$ is a finite set of attributes. A (*fuzzy*) *functional dependence* (*over attributes* $Y$) is an expression $A \Rightarrow B$, where $A, B \in \mathbf{L}^Y$ are fuzzy sets of attributes. We use also "FD" for "functional dependence".

Functional dependencies will be evaluated in the following data tables: A *data table over domains with similarity relations* is a tuple

$$\mathcal{D} = \langle X, Y, \{\langle D_y, \approx_y \rangle \mid y \in Y\}, T \rangle, \text{ where}$$

- $X$ is a non-empty set (of objects, table rows),
- $Y$ is a non-empty finite set (of attributes, table columns),
- 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 in $D_y$ which is reflexive and symmetric (similarity),
- $T$ 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 $x$).

$\mathcal{D}$ always denotes some data table over domains with similarity relations with its components denoted as above, $A \Rightarrow B$ always denotes a FD.

*Remark 1.* (1) $\mathcal{D}$ 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 $D_y$.

(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 $D_y$ ($y \in Y$) [12].

(3) We may assume that attributes from $Y$ are numbered, i.e. $Y = \{y_1, \ldots, y_n\}$. Then, for $x \in X$ and $Z \subseteq Y$, $x[Z]$ denotes a tuple of values $T(x, y)$ for $y \in Z$. For instance, if $Y = \{y_1, \ldots, y_{10}\}$ and $Z = \{y_2, y_3, y_{10}\}$, then $x[Z] = \langle T(x, y_2), T(x, y_3), T(x, y_{10})\rangle$. Moreover, we denote $x[\{y\}]$ by $x[y]$ and identify it with $T(x, y)$.

We want to consider $A \Rightarrow B$ true in $\mathcal{D}$ if "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 $B$". In general, we will consider a degree $a \in L$ to which $A \Rightarrow B$ is true in $\mathcal{D}$, with $a = 1$ meaning that $A \Rightarrow B$ is (fully) true. 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

$$x_1(C) \approx x_2(C) = \bigwedge_{y \in Y} \big(C(y) \to (x_1[y] \approx_y x_2[y])\big). \tag{2}$$

That is, $x_1(C) \approx x_2(C)$ is truth degree of "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 $y$". Then, validity of a FD is 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

$$||A \Rightarrow B||_{\mathcal{D}} = \bigwedge_{x_1, x_2 \in X} \big((x_1(A) \approx x_2(A))^* \to (x_1(B) \approx x_2(B))\big). \tag{3}$$

*Remark 2.* (1) $||A \Rightarrow B||_{\mathcal{D}}$ is a truth degree of "for any objects $x_1, x_2 \in X$: if it is 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 FD. 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 FDs.

(3) We now show that for a FD $A \Rightarrow B$, degrees $A(y) \in L$ and $B(y) \in L$ act 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 $a \to b = 1$ (see [9]), we have

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

Therefore, $||A \Rightarrow B||_{\mathcal{D}} = 1$ means that 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 $y' \in Y$, $B(y') \leq (x_1[y'] \approx_y x_2[y'])$" is (fully) true. 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.

### 3.2 Semantic Entailment

We are going to define the meaning of "$A \Rightarrow B$ follows from a collection $T$ of FDs". Since FDs may be valid to various degrees, we assume that, in general, $T$ encompasses FDs with their degrees of validity. That is, we assume that $T$ is a fuzzy set of FDs and that $T(C \Rightarrow D)$, i.e. degree to which $C \Rightarrow D$ belongs to $T$, is a degree of validity of $C \Rightarrow D$, cf. also [8]. This covers the case when $T$ is crisp (i.e. $T(C \Rightarrow D) = 1$ or $T(C \Rightarrow D) = 0$), i.e. a given FD either is assumed valid or not; then 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 FDs, the set $\mathrm{Mod}(T)$ of all *models* of $T$ is defined by

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

where $\mathcal{D}$ stands for an arbitrary data table over domains with similarities. That is, $\mathcal{D} \in \mathrm{Mod}(T)$ means that for each FD $A \Rightarrow B$, 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 $T$, $\mathrm{Mod}(T) = \{\mathcal{D} \mid \text{for each } A \Rightarrow B \in T\colon ||A \Rightarrow B||_{\mathcal{D}} = 1\}$.

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

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

That is, $||A \Rightarrow B||_T$ is a truth degree of "$A \Rightarrow B$ is true in all models of $T$".

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

$$c \leq ||A \Rightarrow B||_{\mathcal{D}} \quad \textit{iff} \quad ||A \Rightarrow c \otimes B||_{\mathcal{D}} = 1. \tag{4}$$

*Proof.* Sketch: it can be shown that $c \leq ||A \Rightarrow B||_{\mathcal{D}}$ iff $c \to ||A \Rightarrow B||_{\mathcal{D}} = 1$ iff $||A \Rightarrow c \otimes B||_{\mathcal{D}} = 1$.

Lemma 1 enables us to reduce the concept of a model of a fuzzy set of FDs to the concept of a model of an ordinary set of FDs, and to reduce the concept of semantic entailment from a fuzzy set of FDs to the concept of semantic entailment from an ordinary set of FDs:

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

$$\mathrm{c}(T) = \{A \Rightarrow T(A \Rightarrow B) \otimes B \mid A, B \in \mathbf{L}^Y \text{ and } T(A \Rightarrow B) \otimes B \neq \emptyset\}. \quad (5)$$

*Then we have*

$$\mathrm{Mod}(T) = \mathrm{Mod}(\mathrm{c}(T)), \quad (6)$$

$$||A \Rightarrow B||_T = ||A \Rightarrow B||_{\mathrm{c}(T)}, \quad (7)$$

*and thus*

$$||A \Rightarrow B||_T = \bigvee\{c \in L \mid ||A \Rightarrow c \otimes B||_T = 1\}, \quad (8)$$

$$||A \Rightarrow B||_T = \bigvee\{c \in L \mid ||A \Rightarrow c \otimes B||_{\mathrm{c}(T)} = 1\}. \quad (9)$$

*Proof.* Using definitions and Lemma 1.

Note that due to (9), the concept of a degree of entailment from a fuzzy set of FDs can be reduced to entailment in degree 1 from a set of FDs.

## 4 Complete System of Rules for Functional Dependencies

We now introduce an axiomatic system for reasoning with FDs and prove its completeness in two versions. First, we prove that a FD $A \Rightarrow B$ is provable from an ordinary set $T$ of FDs iff $A \Rightarrow B$ semantically follows from $T$ in degree 1 (completeness). Second, we introduce a concept of a degree $|A \Rightarrow B|_T$ of provability of a FD $A \Rightarrow B$ from a fuzzy set $T$ of FDs and show that $|A \Rightarrow B|_T = ||A \Rightarrow B||_T$ (graded completeness, see [8]).

### 4.1 Axioms and Some Derived 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 usual deduction rules: having FDs which are of the form of FDs in the input part (the part preceding "infer") of a rule, a rule allows us to infer the corresponding FD in the output part (the part following "infer") of a rule.

*Remark 3.* (1) Rules (Ax) and (Cut) are taken from [10]. A difference from [10] is that $A, B, C, D$ are fuzzy sets in (Ax) and (Cut) while in [10], $A, B, C, D$ are ordinary sets.

(2) Rule (Mul) is a new rule in our fuzzy setting.

A FD $A \Rightarrow B$ is called *provable* from a set $T$ of FDs, written $T \vdash A \Rightarrow B$, if there is a sequence $\varphi_1, \ldots, \varphi_n$ of FDs such that $\varphi_n$ is $A \Rightarrow B$ and for each $\varphi_i$ we either

have $\varphi_i \in T$ or $\varphi_i$ is inferred (in one step) from some of the preceding FDs (i.e., $\varphi_1, \ldots, \varphi_{i-1}$) using some deduction rule (Ax)–(Mul). A deduction rule "from $\varphi_1, \ldots, \varphi_n$ infer $\varphi$" ($\varphi_i, \varphi$ are FDs) is said to be *derivable* from (Ax)–(Mul) if $\{\varphi_1, \ldots, \varphi_n\} \vdash \varphi$.

**Lemma 3.** *If "from $\varphi_1, \ldots, \varphi_n$ infer $\varphi$" is a rule derivable from the ordinary Armstrong axioms (see [12]) then replacing symbols of sets by symbols of fuzzy sets, the resulting rule is derivable from* (Ax) *and* (Cut).

*Proof.* It follows from [10] that each deduction rule derivable from the ordinary Armstrong axioms is derivable from ($\mathrm{Ax}_c$) and ($\mathrm{Cut}_c$) where ($\mathrm{Ax}_c$) and ($\mathrm{Cut}_c$) result from (Ax) and (Cut) by replacing fuzzy sets by ordinary sets. Now, observe that replacing ordinary sets with fuzzy sets in any proof from ($\mathrm{Ax}_c$) and ($\mathrm{Cut}_c$), we get a proof from (Ax) and (Cut).

*Remark 4.* Lemma 3 shows that, for instance, the following deduction rules are derivable from (Ax) and (Cut):

(Ref) infer $A \Rightarrow A$,
(Wea) from $A \Rightarrow B$ infer $A \cup C \Rightarrow B$,
(Add) from $A \Rightarrow B$ and $A \Rightarrow C$ infer $A \Rightarrow B \cup C$,
(Pro) from $A \Rightarrow B \cup C$ infer $A \Rightarrow B$,
(Tra) from $A \Rightarrow B$ and $B \Rightarrow C$ infer $A \Rightarrow C$,

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

### 4.2 Completeness

A deduction rule "from $\varphi_1, \ldots, \varphi_n$ infer $\varphi$" is said to be *sound* if

$$\mathrm{Mod}(\{\varphi_1, \ldots, \varphi_n\}) \subseteq \mathrm{Mod}(\{\varphi\}),$$

i.e. if each model of all $\varphi_1, \ldots, \varphi_n$ is also a model of $\varphi$.

**Lemma 4.** *Each of the rules* (Ax)–(Mul) *is sound.*

*Proof.* Omitted due to lack of space (proof is straightforward from definitions).

Let $T$ be a set of FDs. $T$ is called *syntactically closed* if $T \vdash A \Rightarrow B$ iff $A \Rightarrow B \in T$, i.e. if $T = \{A \Rightarrow B \,|\, T \vdash A \Rightarrow B\}$. $T$ is called *semantically closed* if $||A \Rightarrow B||_T = 1$ iff $A \Rightarrow B \in T$, i.e. if $T = \{A \Rightarrow B \,|\, ||A \Rightarrow B||_T = 1\}$.

**Lemma 5.** *Let $T$ be a set of FDs. If $T$ is semantically closed then $T$ is syntactically closed.*

*Proof.* Sketch: First it can be shown that a set $T$ of FDs is syntactically closed iff we have: $A \cup B \Rightarrow A \in T$; if $A \Rightarrow B \in T$ and $B \cup C \Rightarrow D \in T$ then $A \cup C \Rightarrow D \in T$; if $A \Rightarrow B \in T$ then $c^* \otimes A \Rightarrow c^* \otimes B \in T$, for each $A, B, C, D \in \mathbf{L}^Y$, and $c \in L$. These conditions are satisfied for if "from $\varphi_1, \ldots, \varphi_n$ infer $\varphi$" is one of (Ax)–(Mul), then if $\varphi_1, \ldots \varphi_n \in T$, we have $\mathrm{Mod}(T) \subseteq \mathrm{Mod}(\{\varphi_1, \ldots \varphi_n\}) \subseteq \mathrm{Mod}(\{\varphi\})$

by soundness of (Ax)–(Mul). This says each model of $T$ is a model of $\varphi$, i.e. $||\varphi||_T = 1$. Since $T$ is semantically closed, i.e. $T = \{A \Rightarrow B \,|\, ||A \Rightarrow B||_T = 1\}$, we get $\varphi \in T$.

**Lemma 6.** *Let $T$ be a set of FDs, 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 $\{A \Rightarrow B \,|\, ||A \Rightarrow B||_T = 1\} \subseteq T$. We prove this by showing that if $A \Rightarrow B \notin T$ then $A \Rightarrow B \notin \{A \Rightarrow B \,|\, ||A \Rightarrow B||_T = 1\}$. By Lemma 4, since $T$ is syntactically closed, it is closed under all of the rules which result from Armstrong axioms (and thus also their consequences) by replacing sets with fuzzy sets. Let thus $A \Rightarrow B \notin T$. To see $A \Rightarrow B \notin \{A \Rightarrow B \,|\, ||A \Rightarrow B||_T = 1\}$, we show that there is $\mathcal{D} \in \mathrm{Mod}(T)$ which is not a model of $A \Rightarrow B$. For this purpose, let first $A^+$ be the largest fuzzy set $C$ such that $A \Rightarrow C \in T$. $A^+$ exists. Namely, $V = \{C \,|\, A \Rightarrow C \in T\}$ is non-empty since $A \Rightarrow A \in T$ by (Ref), $V$ is finite by finiteness of $Y$ and $L$, and for $A \Rightarrow C_1, \ldots, A \Rightarrow C_n \in T$, we have $A \Rightarrow \bigcup_{i=1}^n C_i \in T$ by a repeated use of (Add). Now, take a data table $\mathcal{D}$ with $X = \{x_1, x_2\}$ such that for $y \in Y$ we have: if $A^+(y) = 1$ then $D_y = \{a\}$, $T(x_1, y) = T(x_2, y) = a$, $a \approx_y a = 1$; if $A^+(y) \neq 1$ then $D_y = \{a, b\}$, $T(x_1, y) = a$, $T(x_2, y) = b$, $a \approx_y a = b \approx_y b = 1$, $a \approx_y b = b \approx_y a = A^+(y)$. Then for each $y \in Y$, $\approx_y$ is reflexive and symmetric (and even transitive).

Now, it can be shown that $\mathcal{D}$ is a model of $T$ but not of $A \Rightarrow B$ (details omitted due to lack of space).

We thus have completeness of (Ax)–(Mul).

**Theorem 1 (completeness).** *Let* $\mathbf{L}$ *and $Y$ be finite. Let $T$ be a set of FDs. Then*

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

*Proof.* Sketch: Denote by $syn(T)$ the least syntactically closed set of FDs which contains $T$. It can be shown that $syn(T) = \{A \Rightarrow B \,|\, T \vdash A \Rightarrow B\}$. Furthermore, denote by $sem(T)$ the least semantically closed set of FDs which contains $T$. It can be shown that $sem(T) = \{A \Rightarrow B \,|\, ||A \Rightarrow B||_T = 1\}$. To prove the claim, we need to show $syn(T) = sem(T)$. As $syn(T)$ is syntactically closed, it is also semantically closed by Lemma 6 which means $sem(syn(T)) \subseteq syn(T)$. Therefore, by $T \subseteq syn(T)$ we get

$$sem(T) \subseteq sem(syn(T)) \subseteq syn(T).$$

In a similar manner we get $syn(T) \subseteq sem(T)$, showing $syn(T) = sem(T)$. The proof is complete.

### 4.3 Graded Completeness

Theorem 1 says that for an *ordinary set* $T$ and a FD $A \Rightarrow B$, $A \Rightarrow B$ follows from $T$ *in degree 1* iff $A \Rightarrow B$ is provable from $T$. A question is whether for a *fuzzy*

*set* $T$, a *degree to which* $A \Rightarrow B$ *follows from* $T$ can be somehow approximated using a suitable notion of a proof [8, 9]. In this section, we will see that this is possible, i.e. that (Ax)–(Mul) obey even *completeness in degrees.*

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

$$|A \Rightarrow B|_T = \bigvee\{c \in L \,|\, \mathrm{c}(T) \vdash A \Rightarrow c \otimes B\}, \tag{10}$$

where $\mathrm{c}(T)$ is defined by (5). The following theorem shows that the concept of a degree of provability coincides with that of a degree of semantic entailment.

**Theorem 2 (graded completeness).** *Let* **L** *and* $Y$ *be finite. Then for every fuzzy set* $T$ *of fuzzy attribute implications and* $A \Rightarrow B$ *we have* $|A \Rightarrow B|_T = ||A \Rightarrow B||_T$.

*Proof.* Consequence of Lemma 2 and Theorem 1.

## 5 Computing Non-redundant Bases of All True Functional Dependencies

In the previous sections, we showed that semantic entailment from sets of functional dependencies can be characterized syntactically (by a suitably defined notion of provability / provability degree), i.e. we showed a *completeness* of (Ax)–(Mul). In knowledge engineering, completeness is used still in another sense: "complete" means "fully describing all dependencies which are true in a given data table / model". Therefore, call a set $T$ of functional dependencies *complete in* $\mathcal{D}$ if

$$||A \Rightarrow B||_T = ||A \Rightarrow B||_{\mathcal{D}} \tag{11}$$

for each $A \Rightarrow B$ (degree to which $A \Rightarrow B$ semantically follows from $T$ equals degree to which $A \Rightarrow B$ is true in $\mathcal{D}$). Thus, a set $T$ which is complete in $\mathcal{D}$ conveys all information about dependencies in $\mathcal{D}$ via the concept of semantic entailment. Moreover, if $T$ is complete in $\mathcal{D}$ and no proper subset of $T$ is complete in $\mathcal{D}$, we call $T$ a *non-redundant basis of* $\mathcal{D}$. In other words, a non-redundant basis $T$ is a complete set from which one cannot remove any $A \Rightarrow B \in T$ without losing completeness. From this point of view, we are interested in finding non-redundant bases because they are basically "the minimal sets of FDs conveying the maximal information about $\mathcal{D}$".

Note if $T$ is complete w.r.t. $\mathcal{D}$, it follows immediately from Theorem 1 and the definition of completeness w.r.t. $\mathcal{D}$ that an arbitrary FD $A \Rightarrow B$ can be proved from $T$ using (Ax)–(Mul) iff $A \Rightarrow B$ is true in $\mathcal{D}$ in degree 1.

In the sequel we show a way to compute a non-redundant basis of any $\mathcal{D}$. Since the proofs are technically involved, we omit them due to lack of space.

Given an **L**-set $B$ of attributes, we define a binary **L**-relation $\mathrm{Eq}(B)$ on $X$ (rows of $\mathcal{D}$) as follows

$$(\mathrm{Eq}(B))(x, x') = x(B) \approx x'(B). \tag{12}$$

$\mathrm{Eq}(B)$ is a binary $\mathbf{L}$-relation indicating *similarity of table rows* on attributes from $B$, cf. (2). For any binary $\mathbf{L}$-relation $Sim$ on $X$ we define an $\mathbf{L}$-set $\mathrm{At}(Sim)$ of attributes by

$$(\mathrm{At}(Sim))(y) = \bigwedge_{x,x'} (Sim(x,x') \to (x[y] \approx_y x'[y])). \tag{13}$$

If $Sim$ is interpreted as a similarity relation, $(\mathrm{At}(Sim))(y)$ is a truth degree of "any table rows which are $Sim$-similar are also $Sim$-similar on the value of attribute $y$". Finally, we define an operator $\mathrm{C}\colon \mathbf{L}^Y \to \mathbf{L}^Y$ (i.e., C is an operator on $\mathbf{L}$-sets of attributes) as follows

$$\mathrm{C}(B) = \mathrm{At}((\mathrm{Eq}(B))^*). \tag{14}$$

In words, $(\mathrm{C}(B))(y)$ is a truth degree of proposition: "any table rows which are (very) similar on attributes from $B$ are also similar on the value of attribute $y$". It can be shown that Eq, At, and C given by (12), (13), and (14), respectively, have the following properties (for the notions involved, see e.g. [7]):

**Theorem 3.** Eq *and* At *form a Galois connection.* C *is a closure operator.* □

It can be shown that the set $T = \{B \Rightarrow \mathrm{C}(B) \mid B \in \mathbf{L}^Y\}$ of functional dependencies is complete in $\mathcal{D}$. However, $T$ is not interesting since it is too large and redundant. Nevertheless, $T$ contains non-redundant bases which are based on the following concept.

For any $M \in \mathbf{L}^Y$ (i.e., $M$ is an fuzzy set of attributes) define a data table $\mathcal{D}_M = \langle X, Y, \{\langle D_y, \approx_y\rangle \mid y \in Y\}, T\rangle$ where

- $X = \{x, x'\}$,
- for $y \in Y$, if $M(y) = 1$ then $D_y = \{a\}$, $a \approx_y a = 1$, $T(x,y) = a$, and $T(x',y) = a$,
- for $y \in Y$, if $M(y) \neq 1$ then $D_y = \{a,b\}$, $a \approx_y a = b \approx_y b = 1$, $a \approx_y b = b \approx_y a = M(y)$, $T(x,y) = a$, and $T(x',y) = b$.

Given a data table $\mathcal{D}$ over domains with similarities, $\mathcal{P} \subseteq \mathbf{L}^Y$ (a system of fuzzy sets of attributes) is called a *system of pseudo-intents of* $\mathcal{D}$ if for each $P \in \mathbf{L}^Y$ we have:

$$\begin{aligned} P \in \mathcal{P} \quad \text{iff} \quad & P \neq \mathrm{C}(P) \text{ and } ||Q \Rightarrow \mathrm{C}(Q)||_{\mathcal{D}_P} = 1 \\ & \text{for each } Q \in \mathcal{P} \text{ with } Q \neq P. \end{aligned}$$

The following assertion says that in order to get a non-redundant basis it suffices to pick from $\{B \Rightarrow \mathrm{C}(B) \mid B \in \mathbf{L}^Y\}$ only those FDs where $B$'s belong to a system of pseudo-intents:

**Theorem 4.** *Let* $\mathcal{D} = \langle X, Y, \{\langle D_y, \approx_y\rangle \mid y \in Y\}, T\rangle$ *be a data table over domains with similarities,* $\mathcal{P}$ *be a system of pseudo-intents of* $\mathcal{D}$*. Then* $T = \{P \Rightarrow \mathrm{C}(P) \mid P \in \mathcal{P}\}$ *is a non-redundant basis of* $\mathcal{D}$*.* □

We now show a way to compute a system of pseudo-intents in an efficient way. For brevity, we discuss only particular case for a hedge $^*$ being globalization, i.e.

$a^* = 1$ for $a = 1$ and $a^* = 0$ for $a \neq 1$. First, if $^*$ is globalization then C can be described as follows

$$(\mathrm{C}(B))(y) = \bigwedge\{x[y] \approx_y x'[y] \mid x < x', \text{ and for any } y' \in Y\colon B(y') \leq x[y'] \approx_y x'[y']\}.$$

Furthermore, define an operator $cl_{T^*} : \mathbf{L}^Y \to \mathbf{L}^Y$ (operator on fuzzy sets of attributes) by putting for each $Z \in \mathbf{L}^Y$:

$$Z^{T^*} = Z \cup \bigcup\{B \otimes S(A,Z)^* \mid A \Rightarrow B \in T \text{ and } A \neq Z\},$$
$$Z^{T^*_n} = \begin{cases} Z & \text{if } n = 0, \\ (Z^{T^*_{n-1}})^{T^*} & \text{if } n \geq 1, \end{cases}$$
$$cl_{T^*}(Z) = \bigcup_{n=0}^{\infty} Z^{T^*_n}.$$

The existence and uniqueness of $\mathcal{P}$ is characterized by the following assertion.

**Theorem 5.** *Let* $\mathbf{L}$ *be a finite linearly ordered residuated lattice with globalization,* $\mathcal{D} = \langle X, Y, \{\langle D_y, \approx_y\rangle \mid y \in Y\}, T\rangle$ *be a data table over domains with similarities. Then*

(i) *there is a unique system* $\mathcal{P}$ *of pseudo-intents of* $\mathcal{D}$*;*
(ii) *for* $T = \{P \Rightarrow \mathrm{C}(P) \mid P \in \mathcal{P}\}$, $cl_{T^*}$ *is a closure operator and* $\mathcal{P} \cup \{\mathrm{C}(M) \mid M \in \mathbf{L}^Y\}$ *is the set of all its fixpoints.* □

Hence, in case of globalization and finite linearly ordered structure of truth degrees, one can find $\mathcal{P}$ as a subset of fixpoints of a closure operator. This can be done with polynomial time delay by the following algorithm (inspired by Ganter's NextClosure algorithm [7]):

**Algorithm 1.**

Input: $\mathcal{D}$ (data table over dom. with similarity relations).
Output: $\mathcal{P}$ (system of pseudo-intents).

$B := \emptyset$
*if* $B \neq \mathrm{C}(B)$: *add* $B$ *to* $\mathcal{P}$
*while* $B \neq Y$:
  $T := \{P \Rightarrow \mathrm{C}(P) \mid P \in \mathcal{P}\}$
  $B := B^+$ ($B^+$ is lectically smallest fixed point of $cl_{T^*}$ which is a successor of $B$)
  *if* $B \neq \mathrm{C}(B)$: *add* $B$ *to* $\mathcal{P}$

The efficiency of the previous algorithm depends on computation of $cl_{T^*}(Z)$. A straightforward method to compute $cl_{T^*}(Z)$ leads to an algorithm similar to the CLOSURE algorithm known from database systems [12]. An improved version of CLOSURE, also known as LINCLOSURE [12], can also be adopted in our setting. This and related topics will be discussed in a forthcoming paper.

## 6 Illustrative Examples

Consider again Tab. 1. To get a data table $\mathcal{D} = \langle X, Y, \{\langle D_y, \approx_y \rangle \,|\, y \in Y\}, T \rangle$ over domains with similarity relations, denote by $X$ and $Y$ the sets of planets and their attributes, respectively, put $D_y = [0, \infty)$ for each $y \in Y$, and consider Fig. 1. Fig. 1 depicts similarities on the domains $D_y$. The similarities $\approx_y$ on domains $D_y$ can be described as follows: As a structure of truth degrees, take a real unit interval $[0, 1]$ equipped with Łukasiewicz operations and globalization, and denote by $E_a^b$ a fuzzy set in $[0, \infty)$ defined by

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

$E_a^b$ 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 $E_a^b$. This says that two objects are similar to a degree to which is it true that the objects are "close". Now, the curves depicted in Fig. 1 correspond to similarities defined as follows:

$$x_1 \approx_s x_2 = E_{50}^{500}(|x_1 - x_2|), \quad x_1 \approx_d x_2 = E_{5000}^{20000}(|x_1 - x_2|),$$
$$x_1 \approx_w x_2 = E_1^{10}(|x_1 - x_2|), \quad x_1 \approx_m x_2 = E_1^{5}(|x_1 - x_2|),$$

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. For instance, if $x_1$ denotes *Earth* and $x_2$ denotes *Mars* then "$x_1[m] \approx_m x_2[m] = 1$" (i.e., proposition "Earth and Mars have similar number of moons" is fully true), "$x_1[s] \approx_m x_2[s] \doteq 0.93$" (i.e., proposition "Earth and Mars have similar distance from sun" is true in degree 0.93), etc. Note that "being similar" is subjective and that we can replace the above similarities by other ones.

For technical reasons, we round the exact values of $L = [0, 1]$ from $\approx_y$ ($y \in Y$) down to values of $L = \{0, 0.1, 0.2, \ldots, 0.9, 1\}$. This way we obtain a finite linearly ordered structure of truth degrees with globalization suitable to generate the non-redundant basis of $\mathcal{D}$. In our case, the basis obtained by Algorithm 1 contains the following formulas (for brevity, we do not repeat attributes from premises,

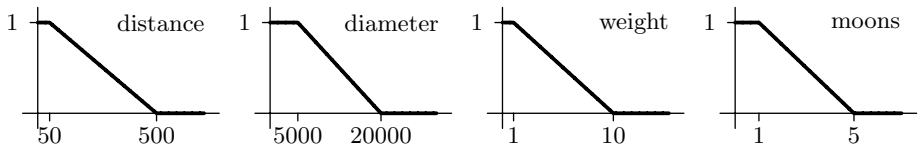


**Fig. 1.** Similarity relations

i.e. instead of FD $A \Rightarrow B$, we list a FD $A \Rightarrow B'$ where $B'$ results from $B$ by deleting all $y$ with $A(y) = B(y)$):

$$\{s, {}^{0.8}\!/d, w, m\} \Rightarrow \{d\}, \qquad \{{}^{0.9}\!/s, {}^{0.8}\!/d, w, {}^{0.7}\!/m\} \Rightarrow \{m\},$$
$$\{{}^{0.8}\!/s, d, w, {}^{0.7}\!/m\} \Rightarrow \{s, m\}, \qquad \{{}^{0.7}\!/s, {}^{0.8}\!/d, w, m\} \Rightarrow \{{}^{0.9}\!/s\},$$
$$\{{}^{0.1}\!/s\} \Rightarrow \{{}^{0.7}\!/s, {}^{0.8}\!/d, w, {}^{0.7}\!/m\}, \qquad \{{}^{0.7}\!/d, {}^{0.8}\!/w\} \Rightarrow \{{}^{0.8}\!/d\},$$
$$\{{}^{0.6}\!/d, w, {}^{0.8}\!/m\} \Rightarrow \{m\}, \qquad \{{}^{0.6}\!/d, {}^{0.9}\!/w\} \Rightarrow \{w, {}^{0.7}\!/m\},$$
$$\{{}^{0.4}\!/d\} \Rightarrow \{{}^{0.6}\!/d, {}^{0.8}\!/w\}, \qquad \{{}^{0.1}\!/d\} \Rightarrow \{{}^{0.3}\!/d\},$$
$$\{{}^{0.1}\!/w\} \Rightarrow \{{}^{0.6}\!/d, {}^{0.8}\!/w\}, \qquad \{{}^{0.1}\!/m\} \Rightarrow \{{}^{0.6}\!/d, w, {}^{0.7}\!/m\}.$$

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

For instance, $\{{}^{0.1}\!/m\} \Rightarrow \{{}^{0.6}\!/d, w, {}^{0.7}\!/m\}$, says "if the numbers of moons are similar in degree (at least) 0.1, then the diameters are similar in degree 0.6, the weights are fully similar, and the numbers of moons are similar in degree 0.7". Taking into account the underlying similarities, the formula can be read:

"if $|x[m] - x[m]| \leq 4$ then $|x[d] - x'[d]| \leq 11000$,
$|x[w] - x'[w]| \leq 1$, and $|x[m] - x'[m] \leq 2|$",

i.e., the implication says: "if the difference between numbers of moons of $x$ and $x'$ is at most 4 then the difference between their diameters is at most 11000, the difference between their weights is at most one weight of Earth, and the difference between numbers of moons is at most 2.

## 7 Concluding Remarks

We introduced functional dependencies for data tables over domains with similarity relations. We presented basic semantic notions (validity, entailment), a complete axiom system, description of non-redundant bases of all functional dependencies which are true in a given table, and presented an algorithm for its computation. In addition to that, in a full version of this paper, we will show

- other complete systems of derivation rules;
- algorithm and related results for other hedges than globalization;
- complete proofs of our theorems.

Note that in a related paper [5] we show a close connection to so-called attribute implications which makes it possible to reduce some problems considered here to analogous problems of fuzzy attribute implications. Our future research will focus on:

- algorithms for various problems of FDs ([12] is a good survey of problems and algorithms in classical FDs);
- further types of data dependencies in a fuzzy setting, like multivalued dependencies (cf. [6]).

## References


1. Abiteboul 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.: *Fuzzy Relational Systems: Foundations and Principles.* Kluwer, Academic/Plenum Publishers, New York, 2002.
4. 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.
5. Bělohlávek R., Vychodil V.: Functional dependencies of data tables over domains with similarity relations (to appear).
6. 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.
7. Ganter B., Wille R.: *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin, 1999.
8. Gerla G.: *Fuzzy Logic. Mathematical Tools for Approximate Reasoning.* Kluwer, Dordrecht, 2001.
9. Hájek P.: *Metamathematics of Fuzzy Logic.* Kluwer, Dordrecht, 1998.
10. Holzer R.: Knowledge Acquisition under Incomplete Knowledge using Methods from Formal Concept Analysis: Part I. *Fundamenta Informaticae,* **63**(1)(2004), 17–39.
11. Klir G. J., Yuan B.: *Fuzzy Sets and Fuzzy Logic. Theory and Applications.* Prentice Hall, 1995.
12. Maier D.: *The Theory of Relational Databases.* Computer Science Press, Rockville, 1983.
13. 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.
14. 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.
15. 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.