

Codd's relational model of data and fuzzy logic: comparisons, observations, and some new results*

Radim Belohlavek and Vilem Vychodil

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

Abstract

The present paper deals with Codd's relational model of data. In particular, we deal with fuzzy logic extensions of the relational model. Our main purpose is to examine relationships between some of the models which have been proposed in the literature. We concentrate on functional dependencies which is the most studied part of the relational model within fuzzy logic extensions. In addition to the observations on relationships between fuzzy logic extensions of the relational model, our paper brings up several new technical results.

1. Introduction and motivation

The foundations of relational databases have been subject to many extensions focusing on issues which, according to the authors of the extensions, are not handled appropriately in the “pure” relational database model. A lot of the extensions comes from the point of view of fuzzy logic. In fact, there is an ongoing stream of papers on fuzzy logic extensions of relational databases starting probably with [11]. Due to lack of space, we omit the discussion on the need to extend current database systems for appropriate treatment of uncertainty and imprecision and refer the reader to many papers and books on this issue. We only note that management of uncertainty in data is listed among six currently most-important research directions proposed in the report from the Lowell debate by 25 senior database researchers [1]. It was pointed out in [1] that “... current DBMS have no facilities for either approximate data or imprecise queries.” Since fuzzy logic and fuzzy set theory offer means to handle particular forms of uncertainty, it is quite natural that fuzzy logic extensions of Codd's relational model have been proposed quite early, more than two decades ago.

Since then, many contributions appeared on fuzzy logic extensions of Codd's model. We found over 100 papers

on fuzzy logic and relational databases (but there are very likely many more papers). The models employ fuzzy logic to enhance various facets of Codd's model in order to make it more appropriate for dealing with uncertainty and imprecision. Two points, which will be discussed in this paper, can be made.

First, the existing approaches are in a sense *ad-hoc*. Namely, the classical Codd's model is strongly connected to first-order logic (predicate logic). The reliance of Codd's model on first-order logic is considered to be the main factor due to which several important issues (data dependence, data integrity, etc.) are successfully handled in Codd's model (see, e.g., [14]). Contrary to that, the link of fuzzy logic extensions of Codd's model to a corresponding logical calculus (i.e., to a fuzzy logic in so-called narrow sense in this case) is missing or, at least, not handled appropriately. On the one hand, this is understandable since foundations of fuzzy logic in narrow sense have been established in the end of 1990s and a solid knowledge of fuzzy logic in narrow-sense is an exception rather than a rule by people who are not directly working in fuzzy logic in narrow sense. On the other hand, it should be clear that the link to an appropriate logical calculus is extremely important particularly in case of fuzzy logic extensions of Codd's model. Namely, when fuzzy logic comes into play (i.e., a scale of truth degrees instead of just 0 and 1, etc.) things become more technically involved. In such a case, a need of a coherent bundle of concepts, rules, and results, such as those provided by fuzzy logic in narrow sense, is apparent. We contend that the lack of foundations of fuzzy logic extensions of Codd's model in fuzzy logic in narrow sense is the main reason why most of the contributions published in the literature are, by and large, “definitional” papers, i.e. papers where results demonstrating feasibility of the introduced concepts are missing. This is perhaps one of the reasons why, up to now, fuzzy logic extensions of Codd's model did not seriously penetrate database community.

Second, there is surprisingly little written about the relationships between various fuzzy logic extensions of Codd's

*Supported by grant No. 1ET101370417 of GA AV ČR, by Kontakt 1–2006–33, and by institutional support, research plan MSM 6198959214.

model and between the concepts investigated within these extensions.

The main aim of this paper is to elaborate on the two points mentioned above. To keep our paper specific and because of the limited scope of our paper, we proceed as follows. First, we describe an extension of Codd’s model and some related notions which we introduced in some of our recent papers, see e.g. [4, 6, 7], and use this extension as a reference model. We pay attention to functional dependencies since they are the most studied concept within fuzzy logic extensions of the relational model. Furthermore, we illustrate a clear connection to fuzzy logic in narrow sense. After that, we comment on several fuzzy logic extensions of the relational model proposed in the literature, discuss some of their mutual relationships, relationships to our model, and present some new results concerning these relationships. Due to the limited scope of this paper, we discuss only selected fuzzy logic extensions and selected aspects. Details including other models and other aspects will be presented in a full version of this paper.

2. Preliminaries

We now recall basic notions of fuzzy logic and fuzzy set theory, for details see, e.g., [3, 15, 16, 20]. We pick so-called complete residuated lattices as our basic structures of truth degrees (i.e., sets of truth degrees equipped with fuzzy logic operations like implication, etc.). Formally, a complete residuated lattice 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: $a \otimes b \leq c$ iff $a \leq b \rightarrow c$ ($a, b, c \in L$). Throughout this paper, \mathbf{L} denotes an arbitrary complete residuated lattice. In addition to that, we consider so-called (truth-stressing) hedges, i.e. unary operations on L satisfying (i) $1^* = 1$, (ii) $a^* \leq a$, (iii) $(a \rightarrow b)^* \leq a^* \rightarrow b^*$, and (iv) $a^{**} = a^*$, for each $a, b \in L$. Elements $a \in L$ are called a truth degrees; \otimes and \rightarrow are (truth functions of) “fuzzy conjunction” and “fuzzy implication”; $*$ is a (truth function of) logical connective “very true” and properties of hedges have natural interpretations, see [16, 17]. 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: Łukasiewicz ($a \otimes b = \max(a + b - 1, 0)$, $a \rightarrow b = \min(1 - a + b, 1)$), Gödel: ($a \otimes b = \min(a, b)$, $a \rightarrow b = 1$ if $a \leq b$, $a \rightarrow b = b$ else), Goguen (product): ($a \otimes b = a \cdot b$, $a \rightarrow b = 1$ if $a \leq b$, $a \rightarrow b = \frac{b}{a}$ else). Complete residuated lattices include also finite structures of truth degrees (e.g., finite Łukasiewicz and Gödel chains). Two boundary cases of hedges are (i) identity, i.e. $a^* = a$ ($a \in L$); (ii) so-called

Table 1. Ranked data table over domains with similarities

$\mathcal{D}(t)$	<i>name</i>	<i>age</i>	<i>education</i>
1.0	Adams	30	Comput. Sci.
1.0	Black	30	Comput. Eng.
0.9	Chang	28	Accounting
0.8	Davis	27	Comput. Eng.
0.4	Enke	36	Electric. Eng.
0.3	Francis	39	Business

$n_1 \approx_n n_2 = \begin{cases} 1 & \text{if } n_1 = n_2 \\ 0 & \text{if } n_1 \neq n_2 \end{cases}$	$a_1 \approx_a a_2 = s_a(a_1 - a_2)$ with scaling $s_a : \mathbb{Z}^+ \rightarrow [0, 1]$	\approx_e	A	B	CE	CS	EE
		A	.7				
		B	.7	1			
		CE			1	.9	.6
		CS			.9	1	.7
		EE			.6	.7	1

globalization: $1^* = 1$, $a^* = 0$ ($a < 1$). 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 $\mathbf{2}$ (structure of truth degrees of classical logic).

An \mathbf{L} -set (a fuzzy set) A in universe U is a mapping $A : U \rightarrow L$, $A(u)$ being interpreted as “the degree to which u belongs to A ”. \mathbf{L}^U denotes the collection of all \mathbf{L} -sets in U . The operations with \mathbf{L} -sets are defined componentwise. For instance, union of \mathbf{L} -sets $A, B \in \mathbf{L}^U$ is an \mathbf{L} -set $A \cup B$ in U such that $(A \cup B)(u) = A(u) \vee B(u)$ ($u \in U$). Binary \mathbf{L} -relations (binary fuzzy relations) in U are just fuzzy sets in $U \times U$.

An \mathbf{L} -set A in U is called crisp if $A(u) = 0$ or $A(u) = 1$ for each $u \in U$. Obviously, crisp \mathbf{L} -sets in U are precisely the characteristic functions of ordinary subsets of U . Therefore, we will identify crisp \mathbf{L} -sets in U with ordinary subsets of U .

3. Extensions of Codd’s model and relationships

3.1. Ranked tables over domains with similarities and their functional dependencies

The main motivation for ranked tables over domains with similarities is the fact that for many domains, it is desirable to consider degrees of similarity of their elements rather than only “equal” and “not equal”. Ranked tables over domains with similarities were introduced in [4, 6, 7]. However, the idea of equipping domains with similarity relations goes back to the early approaches to Codd’s model from the point of view of fuzzy logic, particularly to [11].

The concept of a ranked table over domains with similarities is depicted in Tab. 1. It consists of three parts: data table (relation), domain similarities, and ranking. The data table (right top table in Tab. 1) coincides with a data table of a classical relational model. Domain similarities and rank-

ing are what makes our model an extension of the classical model. The domain similarities (bottom part of Tab. 1) assign degrees of similarity to pairs of values of the respective domain. For instance, a degree of similarity of “Computer Science” and “Computer Engineering” is 0.9 while a degree of similarity of “Computer Science” and “Electrical Engineering” is 0.6. The ranking assigns to each row (tuple) of the data table a degree of a scale bounded by 0 and 1 (left top table in Tab. 1), e.g. 0.9 assigned to the tuple $\langle \text{Chang}, 28, \text{Accounting} \rangle$. The ranking allows us to view the ranked table as an answer to a similarity-based query (rank = degree to which a tuple matches a query). For instance, the ranked table of Tab. 1 can result as an answer to query “show all candidates with age about 30”. In a data table representing stored data (i.e. prior to any querying), ranks of all tuples of the table are equal to 1. Therefore, the same way as tables in the classical relational model, ranked tables represent both stored data and outputs to queries. This is an important feature of our model.

A formal definition follows (Y denotes a set of attributes, attributes are denoted by y, y_1, \dots ; \mathbf{L} denotes a fixed structure of truth degrees).

Definition 1. A ranked data table over domains with similarity relations (with Y and \mathbf{L}) is given by

- *domains*: for each $y \in Y$, D_y is a non-empty set (domain of y , set of values of y);
- *similarities*: for each $y \in Y$, \approx_y is a binary fuzzy relation (called similarity) in D_y (i.e. a mapping $\approx_y: D_y \times D_y \rightarrow L$) which is reflexive (i.e. $u \approx_y u = 1$) and symmetric ($u \approx_y v = v \approx_y u$);
- *ranking*: for each tuple $t \in \times_{y \in Y} D_y$, there is a degree $\mathcal{D}(t) \in L$ (called rank of t in \mathcal{D}) assigned to t .

Remark 2. (1) \mathcal{D} can be seen as a table with rows and columns corresponding to tuples and attributes, like in Tab. 1. By $t[y]$ we denote a value from D_y of tuple t on attribute y . We require that there is only a finite number of tuples which get assigned a non-zero degree (i.e. the corresponding table is finite). Clearly, if $L = \{0, 1\}$ and if each \approx_y is equality, the concept of a ranked data table with similarities coincides with that of a data table (relation) of a classical Codd’s model.

(2) Formally, \mathcal{D} is a fuzzy relation between domains D_y ($y \in Y$). As mentioned above, $\mathcal{D}(t)$ is interpreted as a degree to which the tuple t satisfies requirements posed by a query. We use “non-ranked table” if for each tuple t , $\mathcal{D}(t) = 0$ or $\mathcal{D}(t) = 1$. This accounts for tables representing stored data (prior to querying). Note that ranked tables over domains with similarities appear in [26] which is one of the most advanced and influential papers on fuzzy extensions of Codd’s model. However, the authors consider only $[0, 1]$ as a scale and no logical connectives. Moreover, they do not provide an intuitively clear meaning of the ranking.

(3) One can add additional requirements for \approx_y , e.g. transitivity w.r.t. a binary operation \odot on L , i.e. $(u \approx_y v) \odot (v \approx_y w) \leq (u \approx_y w)$, or separability, i.e. $u \approx_y v = 1$ iff $u = v$, which is sometimes required in the literature. We are not concerned here with how the similarities are represented (we assume they can either be computed or, if D_y is small, are stored).

(4) It is interesting to note that ranked tables over domains with similarities are implicitly used e.g. in [18, 21].

Functional dependencies (FDs) are the most studied data dependencies within fuzzy logic extensions of Codd’s model. In the context of ranked tables over domains with similarities, we introduced FDs as follows [4, 6]:

Definition 3. A (fuzzy) functional dependence is a formula $A \Rightarrow B$ where A and B are fuzzy sets of attributes ($A, B \in \mathbf{L}^Y$).

We first present a definition of validity of $A \Rightarrow B$ in a ranked data table \mathcal{D} and then add comments.

Definition 4. For a ranked data table \mathcal{D} , tuples t_1, t_2 and a fuzzy set $C \in \mathbf{L}^Y$ of attributes, we introduce a degree $t_1(C) \approx_{\mathcal{D}} t_2(C)$ to which t_1 and t_2 have similar values on attributes from C by

$$\begin{aligned} t_1(C) \approx_{\mathcal{D}} t_2(C) &= \\ &= (\mathcal{D}(t_1) \otimes \mathcal{D}(t_2)) \rightarrow \bigwedge_{y \in Y} (C(y) \rightarrow (t_1[y] \approx_y t_2[y])). \end{aligned} \quad (1)$$

A degree $\|A \Rightarrow B\|_{\mathcal{D}}$ to which a FD $A \Rightarrow B$ is true in \mathcal{D} is defined by

$$\begin{aligned} \|A \Rightarrow B\|_{\mathcal{D}} &= \\ &= \bigwedge_{t_1, t_2} ((t_1(A) \approx_{\mathcal{D}} t_2(A))^* \rightarrow (t_1(B) \approx_{\mathcal{D}} t_2(B))). \end{aligned} \quad (2)$$

Remark 5. (1) One can easily see the following: if both A and B are crisp, $A \Rightarrow B$ is just an ordinary FD; if, moreover, each similarity \approx_y is an identity, then $\|A \Rightarrow B\|_{\mathcal{D}} = 1$ iff $A \Rightarrow B$ is true in the ordinary data table corresponding to \mathcal{D} in the usual sense.

(2) By basic rules of semantics of predicate fuzzy logic [16], $t_1(C) \approx_{\mathcal{D}} t_2(C)$ is just the truth degree of a formula “if t_1, t_2 are from \mathcal{D} then for each attribute y from C , t_1 and t_2 have similar values on y ”. Moreover, $\|A \Rightarrow B\|_{\mathcal{D}}$ is a truth degree of a formula “for any tuples t_1, t_2 : if t_1 and t_2 have similar values on attributes from A then t_1 and t_2 have similar values on attributes from B ”. Note that due to our adherence to predicate fuzzy logic, the meaning of $A \Rightarrow B$ is given by a simple formula which we just described in natural language. Note that, in fact, the antecedent in formula (2) is modified by a hedge $*$. This has technical reasons not discussed in detail here (note only that setting $*$ to globalization or identity enables as to have some of the previous approaches as particular cases of our ones).

(3) Note also that $\|A \Rightarrow B\|_{\mathcal{D}}$ is a truth degree from our scale L , not necessarily being 0 or 1, and that this comes up naturally in the context of predicate fuzzy logic. That is, our FDs may be true to a degree, e.g., 0.9 (approximately true) which is natural when considering approximate concepts like similarity. The particular value and the meaning of $\|A \Rightarrow B\|_{\mathcal{D}}$ depends on our choice of the scale and the connectives. For illustration, if the ranks in \mathcal{D} are all 0 or 1 and $*$ is globalization then for any choice of a scale L and connectives \otimes, \rightarrow we have that $\|A \Rightarrow B\|_{\mathcal{D}} = 1$ ($A \Rightarrow B$ is fully true in \mathcal{D}) means that for any tuples t_1, t_2 from \mathcal{D} : if $A(y) \leq (t_1[y] \approx_y t_2[y])$ for any attribute $y \in Y$ then $B(y) \leq (t_1[y] \approx_y t_2[y])$ for any attribute $y \in Y$. This also shows that degrees $A(y)$ and $B(y)$ serve basically as similarity thresholds.

We now turn our attention to a ‘‘classical’’ problem in FDs, namely, axiomatizability by Armstrong axioms. First, we introduce the necessary concepts. For a set T of fuzzy FDs, let $\text{Mod}(T)$ be a set of all ranked data tables with similarities in which each FD from T is true in degree 1, i.e. $\text{Mod}(T) = \{\mathcal{D} \mid \text{for each } A \Rightarrow B \in T : \|A \Rightarrow B\|_{\mathcal{D}} = 1\}$. $\mathcal{D} \in \text{Mod}(T)$ are called models of T . A *degree* $\|A \Rightarrow B\|_T$ to which $A \Rightarrow B$ semantically follows from T is defined by

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

where the infimum ranges over all models of T .

Consider now the following axiomatic system. It 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 $c \in L$. Here, $c^* \otimes A \in \mathbf{L}^Y$ is defined by $(c^* \otimes A)(y) = c^* \otimes A(y)$. As usual, $A \Rightarrow B$ is called *provable* from a set T of FDs, written $T \vdash A \Rightarrow B$, if there is a sequence $\varphi_1, \dots, \varphi_n$ of FDs 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 FDs (i.e., $\varphi_1, \dots, \varphi_{i-1}$) using some deduction rule (Ax)–(Mul).

Remark 6. Note again that the previous concepts are based on the principles of fuzzy logic. As a result, the concepts have a clear meaning. For instance, $\|A \Rightarrow B\|_T$ is a truth degree of a formula saying that $A \Rightarrow B$ is true in each model of T , see [4, 6, 7] for details.

Without going to details, let us note that one can also introduce a *degree* $|A \Rightarrow B|_T \in L$ to which $A \Rightarrow B$ is provable from T by

$$|A \Rightarrow B|_T = \bigvee \{c \in L \mid T \vdash A \Rightarrow c \otimes B\}.$$

The following theorem shows that provability and degree of provability coincide with semantic entailment in degree 1 and degree of entailment, i.e. they are appropriate syntactic notions capturing entailment of fuzzy FDs [6, 7]:

Theorem 7 (completeness and graded completeness). *Let T be a set of FDs, L and Y be finite. For each $A \Rightarrow B$ we have*

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

and

$$|A \Rightarrow B|_T = \|A \Rightarrow B\|_T.$$

Note that the results can be generalized also for fuzzy sets T of formulas (i.e. for reasoning from partially true premises) and that we can also have a so-called Pavelka-style logic for sound and complete reasoning with fuzzy FDs, see [15, 16].

Remark 8. The presented results generalize the ordinary results on completeness of Armstrong axioms [23]. Our results ‘‘become’’ the classical ones if we take a two-element Boolean algebra for our structure of truth degrees.

3.2. Comparisons and observations

We now turn our attention to comparison of selected aspects of fuzzy logic extensions of Codd’s model.

Raju and Majumdar [26] is perhaps the most influential paper on FDs over domains with similarities. Their extension of Codd’s model is a particular case of ranked tables over domains with similarities from Section 3.1 in that they consider only $[0, 1]$ as a structure of truth degrees and they do not consider any (truth function of) logical connective of implication. [26] is probably the first approach considering both ranks and similarities. However, the meaning of ranks is intuitively not very clear. While we interpret a rank assigned to a tuple as a degree to which the tuple matches a similarity-based query (see Section 3.1), Raju and Majumdar describe a rank as a degree to which a tuple belongs to a table. Later on, in Example 3.1, they say that a rank can be interpreted as a possibility measure or a measure of association of the items of a tuple.

Raju and Majumdar consider ordinary FDs in their model, i.e. consider $A \Rightarrow B$ where A and B are crisp sets, and consider a FD $A \Rightarrow B$ true in a ranked table \mathcal{D} if for all tuples t_1, t_2 with $\mathcal{D}(t_1) > 0$ and $\mathcal{D}(t_2) > 0$ we have

$$\min_{y \in Y, A(y)=1} (t_1[y] \approx_y t_2(y)) \leq \min_{y \in Y, B(y)=1} (t_1[y] \approx_y t_2(y)). \quad (3)$$

Consider now the relationship of (3) to $\|A \Rightarrow B\|_{\mathcal{D}}$ from Definition 4. We limit ourselves to the following points. First, [26] consider only ‘‘true’’ and ‘‘not true’’ for a given FD $A \Rightarrow B$. Thus, they disregard possible intermediate truth degrees to which $A \Rightarrow B$ may be true in \mathcal{D} . This may seem as not natural in the context of domain similarities, one might wish to have means to say that $A \Rightarrow B$ is ‘‘almost true’’, i.e.

true in degree e.g. 0.9. Second, the expressive capability of FDs from [26] is smaller than that of our fuzzy FDs from Section 3.1. This is due to the restriction of A and B to be crisp sets. For instance, using FDs of [26], it is not possible to describe a dependence “if similarity of t_1 and t_2 in y is at least 0.5 then similarity of t_1 and t_2 in y' is at least 0.8” (we omit details). Third, the concept $A \Rightarrow B$ of being true in \mathcal{D} according to [26] is a particular case of the concept of $\|A \Rightarrow B\|_{\mathcal{D}}$ in the following sense (we omit proof):

Lemma 9. *For $L = [0, 1]$, A, B crisp, $*$ being identity, and any \rightarrow we have that $A \Rightarrow B$ is true in \mathcal{D} according to (3) iff $\|A \Rightarrow B\|_{\mathcal{D}} = 1$ according to Definition 4.*

Fourth, also with other logical notions like that of semantic consequence, [26] consider these notions as bivalent (e.g., either $A \Rightarrow B$ follows from a set T of FDs or not).

Our last remark concerns Raju and Majumdar’s result on completeness of Armstrong’s axioms (see [23]) w.r.t. to their semantics given by ranked tables with similarities. Raju and Majumdar essentially proved (although they presented their result in a bit different way) that any set \mathcal{R} of deduction rules which is complete w.r.t. ordinary Codd’s model is also complete w.r.t. semantics given by ranked tables with similarities from [26], i.e. a FD $A \Rightarrow B$ semantically follows from a set T of FDs (i.e., $A \Rightarrow B$ is true in each \mathcal{D} such that any FD from T is true in \mathcal{D}) iff $A \Rightarrow B$ can be inferred from T using rules from \mathcal{R} . For this purpose, they elaborated quite a long proof in [26]. Since the completeness holds for any system of ordinary Armstrong deduction rules (i.e., intermediate degrees in a sense do not matter in proofs), one might wonder whether it is possible to obtain the result by some simple reduction to the well-known completeness of the ordinary Codd’s model, see [23]. This is, indeed, the case. Namely, the completeness result of [26] follows almost immediately from the following lemma (we omit proof):

Lemma 10. *A FD $A \Rightarrow B$ follows from a set T of FDs in the sense of [26] (semantics given by ranked tables with similarities) iff $A \Rightarrow B$ follows from T in the sense of ordinary Codd’s model [23].*

This way, we obtain a short proof which, moreover, provides us with an insight about the ordinary semantic consequence and that of [26]. Note also that the completeness result of [26] can also be obtained as a consequence of Theorem 7 (we omit details here). At this point we stop our visit to [26].

Further approaches to FDs over domains with similarities The paper by Raju and Majumdar has been subject to several extensions. We now briefly comment on some of them and leave a more complete discussion to a full version of the paper (where we will discuss, among others,

also [10, 19, 22, 27, 28, 29]). In all of the subsequent approaches, a FD is considered as a formula $A \Rightarrow B$ where A and B are crisp (i.e. $A \Rightarrow B$ is an ordinary FD) possibly with additional parameters, and FDs are being evaluated in data tables \mathcal{D} over domains with similarities (without ranks). In addition to that, L is always confined to $[0, 1]$. In [13], FDs are parameterized by $c_y \in [0, 1]$ ($y \in Y$). Values c_y are fixed and common to any FDs considered. A FD $A \Rightarrow B$ (denoted by the authors by $A \Rightarrow_{(\alpha, \beta)} B$ with $\alpha = (c_y)_{y \in A}$ and $\beta = (c_y)_{y \in B}$) is considered true in \mathcal{D} whenever: if for each $y \in A$ we have $t_1[y] \approx_y t_2[y] \geq c_y$, then for each $y \in B$ we have $t_1[y] \approx_y t_2[y] \geq c_y$. One can see that if we define fuzzy sets A_c and B_c by $A_c(y) = c_y$ for $y \in A$ and $A_c(y) = 0$ for $y \notin A$ (and the same for B_c), we have (proof omitted):

Lemma 11. *For $*$ being globalization, $L = [0, 1]$, and any \rightarrow , $A \Rightarrow B$ is true in \mathcal{D} according to [13] iff $\|A_c \Rightarrow B_c\|_{\mathcal{D}} = 1$, cf. (2).*

Therefore, [13] results as a particular instance of our above approach (which, moreover, does not require fixed thresholds c_y). In [9], a FD $A \Rightarrow B$ is considered true in degree b in \mathcal{D} if $b = \min_{t_1, t_2} (\min(1, 1 - (t_1(A) \approx_{\mathcal{D}} t_2(A) + (t_1(B) \approx_{\mathcal{D}} t_2(B))))$ and if for all tuples t_1, t_2 we have $\min(1, 1 - (t_1(A) \approx_{\mathcal{D}} t_2(A) + (t_1(B) \approx_{\mathcal{D}} t_2(B))) \geq \theta$ ($\theta \in [0, 1]$ is an additional parameter). The basic relationship to our approach is the following (proof omitted):

Lemma 12. *For $*$ being identity, $L = [0, 1]$, and Lukasiewicz \rightarrow , $A \Rightarrow B$ is true in degree b in \mathcal{D} according to [9] iff $b = \|A \Rightarrow B\|_{\mathcal{D}}$ and $b \geq \theta$, cf. (2).*

A similar approach is adopted in [8]. In [12], a FD $A \Rightarrow_{\gamma} B$ (with $\gamma \in [0, 1]$ a parameter) is considered true in \mathcal{D} if $A \Rightarrow B$ is true in \mathcal{D} in the ordinary sense and if

$$t_1(A) \approx_{\mathcal{D}} t_2(A) \rightarrow t_1(B) \approx_{\mathcal{D}} t_2(B) \geq \gamma$$

where \rightarrow is Gödel implication, cf. Section 2. The following relationship to our model is almost obvious:

Lemma 13. *For $*$ being identity, $L = [0, 1]$, and Gödel \rightarrow , $A \Rightarrow_{\gamma} B$ is true in \mathcal{D} according to [12] iff $\|A \Rightarrow B\|_{\mathcal{D}} \geq \gamma$, cf. (2), and if $A \Rightarrow B$ is true in \mathcal{D} as an ordinary FD.*

Prade and Testemale and related approaches [25] is a seminal paper on another extension of the relational model from the point of view of fuzzy logic. Due to lack of space we only briefly comment on it. The main idea of this extension consists in allowing fuzzy sets A_y in domains D_y as members of the tuples. FDs in this model are based on similarity relations between fuzzy sets A_y and A'_y . An important observation here is that, for treatment of some aspects, one can consider the FDs in this model as a particular case of FDs in (ranked) tables over domains with similarities. Namely, the elements of the domains are fuzzy sets A_y

and the similarities on the domains are the above-mentioned similarity relations between fuzzy sets. Then, one can employ results on FDs of tables over domains with similarities to the model of [25] and to related models.

4. Conclusions and further issues

We have demonstrated some of the benefits of developing fuzzy logic extensions of Codd's relational model according to the principles of fuzzy logic in narrow sense by presenting an example of such an extension. Furthermore, we compared some of the influential approaches from the literature with the presented extension and derived some new observations and results. The main conclusion is that not only are several of the approaches proposed in the literature particular instances of our extension (with possibly imposing additional conditions in the definition of validity) but, more importantly, our extension, based on the principles of fuzzy logic, is transparent and tractable from both theoretical and algorithmic point of view.

In addition to a more thorough treatment of the above topics, a full version of this paper will include relational algebra and calculus (this is an interesting part of the fuzzy logic extensions since it naturally includes operations/queries like “find tuples with age similar to 35” but also “find top k tuples with age similar to 30 and education similar to Computer Science” which were recently discussed in database community, see [7]), algorithmic aspects (for instance, there are several result available for computation of various problems related to semantic entailment of FDs in fuzzy setting), and a more thorough discussion on Armstrong-like axiomatization of FDs.

References

- [1] S. Abiteboul *et al.* The Lowell database research self-assessment. *Comm. ACM* **48**(5)(2005), 111–118.
- [2] W. W. Armstrong. Dependency structures in data base relationships. *IFIP Congress*, Geneva, Switzerland, 1974, pp. 580–583.
- [3] R. Belohlavek. *Fuzzy Relational Systems: Foundations and Principles*. Kluwer, Academic/Plenum Publishers, New York, 2002.
- [4] R. Belohlavek and V. Vychodil. Functional dependencies of data tables over domains with similarity relations. In: *Proc. IICAI 2005*, pp. 2486–2504, ISBN 0–9727412–1–6.
- [5] R. Belohlavek and V. Vychodil. Attribute implications in a fuzzy setting. In: Missaoui R., Schmid J. (Eds.): *ICFCA 2006*, LNAI **3874**, pp. 45–60, 2006.
- [6] R. Belohlavek and V. Vychodil. Data tables with similarity relations: functional dependencies, complete rules and non-redundant bases. In: M.L. Lee, K.L. Tan, and V. Wuwongse (Eds.): *DASFAA 2006*, LNCS **3882**, pp. 644–658, Springer-Verlag, Berlin/Heidelberg, 2006.
- [7] R. Belohlavek and V. Vychodil. Relational Model of Data over Domains with Similarities: An Extension for Similarity Queries and Knowledge Extraction. *IEEE IRI 2006*, Sep 16–18, 2006, Waikoloa, Hawaii, USA (to appear).
- [8] B. Bhuniya, P. Niyogi: Lossless join property in fuzzy relational databases. *Data Knowledge Eng.* **11**(2)(1993), 109–124.
- [9] S. Ben Yahia, H. Ounalli, and A. Jaoua. An extension of classical functional dependency: dynamic fuzzy functional dependency. *Inf. Sci.* 119:219–234, 1999.
- [10] P. Bosc, D. Dubois, and H. Prade. Fuzzy functional dependencies. An overview and a critical discussion. *FUZZ-IEEE'94*, pages 325–330, Orlando, FL, 1994.
- [11] B. P. Buckles and F. E. Petry. A fuzzy representation of data for relational databases. *Fuzzy Sets Syst.* 7:213–226, 1982.
- [12] G. Chen, E. E. Kerre, and J. Vandenbulcke. A computational algorithm for the FFD transitive closure and a complete axiomatization of fuzzy functional dependence (FFD). *Int. J. Intel. Systems*, Vol. **9**(1994), 421–439.
- [13] J. C. Cubero and M. A. Vila. A new definition of fuzzy functional dependency in fuzzy relational datatbases. *Int. J. Intelligent Systems* 9:441–448, 1994.
- [14] C. J. Date. *Database Relational Model: A Retrospective Review and Analysis*. Addison Wesley, 2000.
- [15] G. Gerla. *Fuzzy Logic. Mathematical Tools for Approximate Reasoning*. Kluwer, Dordrecht, 2001.
- [16] P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer, Dordrecht, 1998.
- [17] P. Hájek. On very true. *Fuzzy Sets and Syst.* 124:329–333, 2001.
- [18] I. F. Ilyas, W. G. Aref, and A. K. Elmagarmid. Supporting top- k join queries in relational databases. *The VLDB Journal* 13:207–221, 2004.
- [19] R. Intan and M. Mukaidono. Fuzzy conditional probability relations and their applications in fuzzy information systems. *Knowledge and Inf. Systems* 6:345–365, 2004.
- [20] G. J. Klir and B. Yuan. *Fuzzy Sets and Fuzzy Logic. Theory and Applications*. Prentice Hall, 1995.
- [21] C. Li, K. C.-C. Chang, I. F. Ilyas, and S. Song. RanSQL: Query Algebra and Optimization for Relational top- k queries. *ACM SIGMOD 2005*, pages 131–142, 2005.
- [22] W.-Y. Liu: Fuzzy data dependencies and implication of fuzzy data dependencies. *Fuzzy Sets Syst.* **92**(1997), 341–348.
- [23] D. Maier. *The Theory of Relational Databases*. Computer Science Press, Rockville, 1983.
- [24] F. Petry. *Fuzzy Databases: Principles and Applications*. Kluwer Academic, 1996.
- [25] H. Prade and C. Testemale. Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries. *Inf. Sci.* 34:115–143, 1984.
- [26] K. V. S. V. N. Raju, and A. K. Majumdar. Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems. *ACM Trans. Database Systems* Vol. 13, No. 2:129–166, 1988.
- [27] S. Sheno and A. Melton: Functional dependencies and normal forms in the fuzzy relational database model. *Inf. Sci.* **100**(1992), 1–28.
- [28] B. K. Tyagi, A. Sharfuddin, R. N. Dutta, D. K. Tayal. A complete axiomatization of fuzzy functional dependencies using fuzzy function. *Fuzzy Sets and Syst.* 151:363–379, 2005.
- [29] S.-L. Wang, J.-S. Tsai, and T.-P. Hong. Mining Functional Dependencies from Fuzzy Relational Databases. *ACM SAC 2000*, pages 490–493, Como, Italy, March 2000.