Journal of Logic and Computation Advance Access published September 14, 2009

Codd's Relational Model from the Point of View of Fuzzy Logic

RADIM BELOHLAVEK, *State University of New York at Binghamton, Binghamton, NY 13902, USA*
E-mail: radim.belohlavek@upol.cz

VILEM VYCHODIL, *Palacky University, Olomouc, Tomkova 40, CZ-779 00 Olomouc, Czech Republic.*
E-mail: vilem.vychodil@upol.cz

Abstract

The article deals with Codd's relational model of data and its fuzzy logic extensions. Our main purpose is to examine, from the point of view of fuzzy logic in the narrow sense, some of the extensions proposed in the literature and the relationships between them. We argue that fuzzy logic in the narrow sense is important for the fuzzy logic extensions because it provides conceptual and methodological foundations, clarity and simplicity. We present several comparative observations as well as new technical results.

Keywords: Relational database, relational database model, functional dependency, fuzzy logic

1 Introduction

Codd's relational model of data [16], which is the core of relational databases, has been subject to many extensions. These extensions focus on issues which are not handled appropriately in the ordinary Codd's relational model. Many of the extensions come from fuzzy logic. There is an ongoing stream of papers on fuzzy logic extensions of relational databases starting probably with [13], which address various issues regarding management of uncertainty and imprecision in databases. The need for this line of research is well documented by a report from the Lowell debate by senior database researchers [1], in which management of uncertainty in data is listed among six currently most important research directions. It was pointed out in [1] that '… current DBMS have no facilities for either approximate data or imprecise queries.'

It has been repeatedly argued in the literature that one of the key factors of success of the ordinary Codd's model of data is its foundation in and a clear connection to mathematical logic: 'The relational approach really is rock solid, owing (once again) to its basis in mathematics and predicate logic.' [16, p. 138]. The situation is similar in the case of fuzzy logic extensions of Codd's relational model. A solid approach needs to be based on a solid logic framework, in a clear and simple way. We argue in this article that fuzzy logic in the narrow sense provides such a framework. In addition, we examine several approaches which appeared in the literature and present some of their properties and mutual relationships from a logic point of view. We focus primarily on functional dependencies (FDs) since they represent a part of Codd's model which has been discussed in the literature to the greatest extent.

We found over 100 papers on fuzzy logic and relational databases. Two points, which are discussed in this article, can be made. First, as a rule, the existing approaches are *ad hoc* in the sense that a clear connection to a corresponding logic framework is missing or not handled appropriately. Note

doi:10.1093/logcom/exp056

that this is understandable since foundations of fuzzy logic in the narrow sense were generally not available at the time these approaches were published. It should be clear, however, that the link to an appropriate logical framework is extremely important 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 case, a need of a coherent bundle of concepts, rules and results, such as those provided by fuzzy logic in the narrow sense, is apparent. We contend that the lack of foundations of fuzzy logic extensions of Codd's model in fuzzy logic in the 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 model and between the concepts investigated within these extensions.

This article is organized as follows. Section 2 presents preliminaries from fuzzy logic and fuzzy sets. Section 3 presents an extension of Codd's model which is based on the framework of fuzzy logic in the narrow sense. In Section 4, we examine selected fuzzy logic extensions of Codd's model proposed in the literature. Section 5 contains conclusions and directions for further research.

2 Preliminaries

We now recall basic notions of fuzzy logic and fuzzy set theory. For details see, e.g. [3, 17–19]. We use complete residuated lattices as our basic structures of truth degrees. Recall that 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 the adjointness property: $a \otimes b \leq c$ iff $a \leq b \rightarrow c$ $(a, b, c \in L)$. Throughout this article, $\mathbf{L}$ denotes an arbitrary complete residuated lattice. In addition to that, we consider (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 is very close to the connective of hedge used in [19, 20]. 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) and Goguen (product): ($a \otimes b = a \cdot b$, $a \rightarrow b = 1$ if $a \leq b$, $a \rightarrow b = \frac{b}{a}$ else). Two boundary cases of hedges are (i) identity, i.e. $a^* = a$ $(a \in L)$; (ii) globalization [32]: $1^* = 1$, $a^* = 0$ for $a < 1$. 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 (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. As usual, we identify crisp $\mathbf{L}$-sets in U with ordinary subsets of U.

3 Ranked tables over domains with similarities and their FDs

In this section, we present an extension of Codd's relational model of data, studied in [4, 6, 7]. We focus on its part related to FDs and demonstrate a usefulness of the framework of fuzzy logic in the narrow sense, in particular of Pavelka's abstract fuzzy logic [17, 19, 27], cf. also [26].

We start with the concept of a ranked table over domains with similarities. This is a counterpart to the concept of a table (relation) of the ordinary Codd's model. The motivation for this concept 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 [13].

The concept of a ranked table over domains with similarities is depicted in Table 1. It consists of three parts: data table (relation), domain similarities, and ranking. The data table (right top table in Table 1) coincides with a data table of a classical relational model. Domain similarities and ranking are what makes this model an extension of the ordinary model. The domain similarities (bottom part of Table 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 Table 1), e.g. 0.9 assigned to the tuple $\langle \text{Chang}, 28, \text{Accounting} \rangle$. The ranking allows us to regard a ranked table as an answer to a similarity-based query. The idea is that the degree associated with a given tuple (rank) is the degree to which the tuple matches the query. For instance, the ranked table in Table 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 ordinary relational model, ranked tables represent both stored data and outputs to queries. This is an important feature of the model.

A formal definition follows (Y denotes a set of attributes, attributes are denoted by $y, y_1, \ldots$; $\mathbf{L}$ 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

TABLE 1. Ranked data table over domains with similarities

$\mathcal{D}(t)$	*name*	*age*	*education*
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

$$n_1 \approx_n n_2 = \begin{cases} 1 & \text{if } n_1 = n_2 \\ 0 & \text{if } n_1 \neq n_2 \end{cases}$$

$$a_1 \approx_a a_2 = s_a(|a_1 - a_2|) \quad \text{with scaling } s_a : \mathbb{Z}^+ \to [0,1]$$

$\approx_e$	A	B	CE	CS	EE
A	1	.7			
B	.7	1			
CE			1	.9	.6
CS			.9	1	.7
EE			.6	.7	1

- *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 $\approx_y: D_y \times D_y \to L$), 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 Table 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 the ordinary 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 [30] which is one of the most advanced and influential papers on fuzzy logic 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) From the point of view of logic, ranked tables with similarities are (semantic) structures.

(4) One can add additional requirements for $\approx_y$, e.g. transitivity w.r.t. a binary operation $\odot$ on L, i.e. $(u \approx_y v) \odot (v \approx_y w) \le (u \approx_y w)$, 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 that they can either be computed or, if D_y is small, are stored). Note also that the term tolerance is usually used for reflexive and symmetric relations.

(5) It is interesting to note that ranked tables over domains with similarities are implicitly used, e.g. in [21, 23].

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 an expression $A \Rightarrow B$ where A and B are fuzzy sets of attributes ($A, B \in \mathbf{L}^Y$).

Remark 4

From the point of view of logic, fuzzy FDs are formulas. In Pavelka's abstract fuzzy logic [19, 27], formulas are elements of an (abstract) set. In our case, such set can be thought of as consisting of all expressions $A \Rightarrow B$ (or pairs $\langle A, B \rangle$) of fuzzy sets A and B. Note also that fuzzy FDs can be understood as (abbreviations for) formulas in a first-order fuzzy logic, which in particular contains a unary connective of very true and truth constants in language, see Definition 5.

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

Definition 5

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

$$t_1(C) \approx_{\mathcal{D}} t_2(C) = (\mathcal{D}(t_1) \otimes \mathcal{D}(t_2)) \to \textstyle\bigwedge_{y \in Y} (C(y) \to (t_1[y] \approx_y t_2[y])). \quad (1)$$

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

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

REMARK 6

(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 ordinary sense.

(2) By basic rules of fuzzy logic, $t_1(C) \approx_{\mathcal{D}} t_2(C)$ is just the truth degree of '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 '*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*'. That is, 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 *. The reason for this is that such a parameterized approach (* as a parameter) covers two important special cases which correspond to * being globalization and * being identity, see e.g. [6], which can thus be handled simultaneously using the parameterized approach. Note also that setting * to globalization or identity enables us to regard some of the approaches which appeared in the literature as particular cases of our approach.

(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 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 classic problem in FDs, namely, axiomatizability by Armstrong axioms. First, we introduce the necessary concepts. For this purpose, we can utilize Pavelka's abstract fuzzy logic. Namely, we have FDs as formulas, ranked tables as structures and a rule assigning a truth degree to which a formula is true in a structure. Pavelka's abstract logic gives us immediately the following concepts. For a set T of fuzzy FDs, let $\mathrm{Mod}(T)$ be a set of all ranked data tables with similarities in which each FD from T is true to degree 1, i.e.

$$\mathrm{Mod}(T) = \{\mathcal{D} \mid \text{ for each } A \Rightarrow B \in T : ||A \Rightarrow B||_{\mathcal{D}} = 1\}.$$

$\mathcal{D} \in \mathrm{Mod}(T)$ are called models of T. A *degree* $||A \Rightarrow B||_T$ *to which* $A \Rightarrow B$ *semantically follows from* T is defined by

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

Consider now the following axiomatic system. It consists of three 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 $\varphi_1, \ldots, \varphi_n$ of

FDs such that φ_n is $A \Rightarrow B$ and for each φ_i, either $\varphi_i \in T$ or φ_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).

REMARK 7
The above concepts have a clear meaning, well understood in fuzzy logic. 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 [4–7] for details.

Note that one can also consider theories being fuzzy sets of FDs and the corresponding notion of proof and degree of provability [17, 19, 27]. This is discussed in [8].

In order to syntactically grasp semantic entailment to arbitrary degree using the above introduced concept of a proof, one can 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 to degree 1 and degree of entailment, i.e. they are appropriate syntactic notions capturing entailment of fuzzy FDs [6, 7]:

THEOREM 8 (completeness and graded completeness)
Let T be a set of FDs, $\mathbf{L}$ 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

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

REMARK 9
The presented results generalize the ordinary results on completeness of Armstrong axioms [25]. Our results 'become' the ordinary ones if we take the two-element Boolean algebra for our structure of truth degrees. Note also that a completeness theorem for infinite $\mathbf{L}$ is shown in [9].

4 Examination of selected extensions of Codd's model

We now examine selected fuzzy logic extensions of Codd's model proposed in the literature. We discuss in detail the model proposed by Raju and Majumdar and then briefly comment on some of the further approaches.

4.1 Raju and Majumdar's Model

The paper by Raju and Majumdar [30] is perhaps the most influential one on FDs over domains with similarities. Raju and Majumdar's extension of Codd's model is a particular case of a ranked table over domains with similarities from Section 3 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. The paper by Raju and Majumdar [30] 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), Raju and

Majumdar describe a rank as a possibility measure or a measure of association of the items of a tuple [30, Example 3.1]. In addition, the ranks play no role for FDs in Raju and Majumdar's model.

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

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

Consider now the relationship of (3) to $||A \Rightarrow B||_{\mathcal{D}}$ from Definition 5. We limit ourselves to the following points.

First, [30] 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 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 to degree, e.g. 0.9.

Second, the expressive capability of FDs from [30] is smaller than that of our fuzzy FDs from Section 3. This is due to the restriction of A and B to crisp sets. For instance, using FDs of [30], 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'.

Third, the concept $A \Rightarrow B$ of being true in $\mathcal{D}$ according to [30] is a particular case of the concept of $||A \Rightarrow B||_{\mathcal{D}}$ in the following sense.

Lemma 10
Denote by $\mathcal{D}'$ the strong 0-cut of $\mathcal{D}$, i.e. $\mathcal{D}'(t) = 1$ iff $\mathcal{D}(t) > 0$. 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 5.

Proof. Follows easily from the definitions using the fact that $a \rightarrow b = 1$ if and only if $a \leq b$, for any $a, b \in L$. ■

Fourth, in general, the authors in [30] consider various logical notions, such as that of semantic consequence, bivalent notions (e.g. either $A \Rightarrow B$ follows from a set T of FDs or not), while in our approach, these notions come in degrees.

Our last remark concerns Raju and Majumdar's result on completeness of Armstrong's axioms [25] w.r.t. 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 [30], i.e. a FD $A \Rightarrow B$ semantically follows from a set T of FDs (i.e. $A \Rightarrow B$ is true in every $\mathcal{D}$ in which each FD from T is true) iff $A \Rightarrow B$ can be inferred from T using rules from $\mathcal{R}$. For this purpose, they elaborated quite a long proof in [30]. 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 a simple reduction to the well-known completeness of the ordinary Codd's model, [25]. This is, indeed, the case. Namely, the completeness result of [30] follows immediately from the following lemma.

Lemma 11
A FD $A \Rightarrow B$ follows from a set T of FDs in the sense of [30] (semantics given by ranked tables with similarities) iff $A \Rightarrow B$ follows from T in the sense of ordinary Codd's model [25].

The proof of Lemma 11 makes use of the concept of validity of an ordinary FD in a binary table. Each such table can be identified with a binary relation $I \subseteq X \times Y$ where X and Y contain elements which correspond to the rows and columns of the table, and $\langle x,y\rangle \in I$ iff the table contains 1 in the entry corresponding to row x and column y. Then, given a FD $A \Rightarrow B$, we put $||A \Rightarrow B||_I = 1$ ($A \Rightarrow B$ is true in I) if for every $x \in X$: if $\langle x,y\rangle \in I$ for every $y \in A$ then $\langle x,y\rangle \in I$ for every $y \in B$; and put $||A \Rightarrow B||_I = 0$ otherwise. We now turn to the proof of Lemma 11.

PROOF OF LEMMA 11. Denote by $||A \Rightarrow B||^{\mathrm{RM}}_{\mathcal{D}}$ the truth degree of $A \Rightarrow B$ in a ranked table with similarities according to (3); denote by $||A \Rightarrow B||^{\mathrm{bin}}_I$ the truth degree of $A \Rightarrow B$ in a binary relation, as introduced before this proof. We show that a FD $A \Rightarrow B$ follows from a set T of FDs w.r.t. semantics given by (3) iff $A \Rightarrow B$ follows from T w.r.t. the semantics given by binary tables (binary relations). One can easily check that to prove this, it is sufficient to verify that (A) for every I there exists a ranked table with similarities $\mathcal{D}$ such that $||A \Rightarrow B||^{\mathrm{bin}}_I = ||A \Rightarrow B||^{\mathrm{RM}}_{\mathcal{D}}$, and that (B) for every ranked table with similarities $\mathcal{D}$ there exists a binary relation I such that $||A \Rightarrow B||^{\mathrm{bin}}_I = ||A \Rightarrow B||^{\mathrm{RM}}_{\mathcal{D}}$.

The proof of (A) is easy: it is well known [25] that for I there exists an ordinary table $\mathcal{D}$ such that $||A \Rightarrow B||^{\mathrm{bin}}_I$ equals the truth value of $A \Rightarrow B$ in $\mathcal{D}$. Since the ordinary table $\mathcal{D}$ can be considered a ranked table with similarities (ranks are 1s, and similarities are identity relations) and since (3) coincides with the ordinary concept of validity of FDs in this case, (A) readily follows.

To prove (B), consider $\mathcal{D}$ and denote by Z the set of all tuples t for which $\mathcal{D}(t) > 0$. Furthermore, for $C \subseteq Y$, put

$$t_1[C] \approx t_2[C] = \textstyle\bigwedge_{y \in C} t_1[y] \approx_y t_2[y].$$

Now, let $X = Z \times Z \times L$ and consider a binary relation $I \subseteq X \times Y$ defined by

$$\langle\langle t_1, t_2, a\rangle, y\rangle \in I \text{ iff } (t_1[y] \approx_y t_2[y]) \geq a.$$

One can now verify that $||A \Rightarrow B||^{\mathrm{bin}}_I = 1$ iff for every $\langle t_1, t_2, a\rangle$,

$$\text{if } (t_1[y] \approx_y t_2[y]) \geq a \text{ for each } y \in A, \text{ then } (t_1[y] \approx_y t_2[y]) \geq a \text{ for each } y \in B,$$

i.e. iff for every $\langle t_1, t_2, a\rangle$,

$$\text{if } (t_1[A] \approx t_2[A]) \geq a, \text{ then } (t_1[B] \approx t_2[B]) \geq a,$$

which holds true iff for every t_1, t_2,

$$(t_1[A] \approx t_2[A]) \leq (t_1[B] \approx t_2[B]),$$

i.e. iff $||A \Rightarrow B||^{\mathrm{RM}}_{\mathcal{D}} = 1$, finishing the proof. ■

This way, we obtain a short, reduction-to-ordinary-case proof of the afore mentioned completeness theorem from [30].

THEOREM 12
Let $\mathcal{R}$ be a set of deduction rules for FDs which is syntactico-semantically complete w.r.t. the ordinary semantics of FDs. Then $\mathcal{R}$ is syntactico-semantically complete w.r.t. the semantics of FDs given by (3).

PROOF. A direct consequence of Lemma 11. ■

The proof also provides us with an insight regarding the relationship between the ordinary semantic consequence and that of [30]. Note also that the completeness result of [30] can also be obtained as a consequence of Theorem 8 (we omit details here). At this point we stop our visit to [30].

4.2 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.

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 evaluated in data tables $\mathcal{D}$ over domains with similarities (without ranks). In addition to that, L is always confined to $[0,1]$. In [15], FDs are parameterized by $c_y \in [0,1]$ $(y \in Y)$. Values c_y are fixed and common to all 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:

$$\begin{aligned}&\text{if for each } y \in A \text{ we have } t_1[y] \approx_y t_2[y] \geq c_y\\&\text{then for each } y \in B \text{ we have } t_1[y] \approx_y t_2[y] \geq c_y.\end{aligned}$$

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:

Lemma 13
For * being globalization, $L = [0,1]$, and any $\rightarrow$, $A \Rightarrow B$ is true in $\mathcal{D}$ according to [15] iff $||A_c \Rightarrow B_c||_{\mathcal{D}} = 1$, cf. (2).

Proof. The proof follows from definitions observing that if * is globalization then $(a \rightarrow b)^* = 1$ iff $a \leq b$. ■

Therefore, [15] results as a particular instance of our above approach (which, moreover, does not require fixed thresholds c_y). In [11], a FD $A \Rightarrow B$ is considered true to degree b in $\mathcal{D}$ if

$$\begin{aligned}&b = \min_{t_1,t_2}(\min(1, 1-(t_1(A) \approx_{\mathcal{D}} t_2(A)) + (t_1(B) \approx_{\mathcal{D}} t_2(B)))), \text{ and}\\&\text{for all } t_1, t_2 \text{ we have } \min(1, 1-(t_1(A) \approx_{\mathcal{D}} t_2(A)) + (t_1(B) \approx_{\mathcal{D}} t_2(B))) \geq \theta,\end{aligned}$$

where $\theta \in [0,1]$ is an additional parameter. The basic relationship to our approach is the following:

Lemma 14
For * being identity, $L = [0,1]$, and Łukasiewicz $\rightarrow$, $A \Rightarrow B$ is true to degree b in $\mathcal{D}$ according to [11] iff $b = ||A \Rightarrow B||_{\mathcal{D}}$ and $b \geq \theta$, cf. (2).

Proof. By direct calculation using the fact that $a \rightarrow b = \min(1, 1-a+b)$ for Łukasiewicz $\rightarrow$. ■

A similar approach is adopted in [10]. In [14], 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 15
For * being identity, $L = [0,1]$, and Gödel $\rightarrow$, $A \Rightarrow_\gamma B$ is true in $\mathcal{D}$ according to [14] iff $||A \Rightarrow B||_{\mathcal{D}} \geq \gamma$, cf. (2), and if $A \Rightarrow B$ is true in $\mathcal{D}$ as an ordinary FD.

4.3 Prade and Testemale and related approaches

The paper by Prade and Testemale [29] is a seminal paper on another extension of the relational model from the point of view of fuzzy logic. We briefly mention a basic connection to the model presented in Section 3. The main idea of this extension consists in allowing fuzzy sets in domains D_y as members of the tuples, i.e. $t[y]$ are fuzzy sets. FDs in this model are based on similarity relations between these fuzzy sets. Using our notation, the authors in [29] say that a FD $\{y_1\} \Rightarrow \{y_2\}$ is true if and only if for every tuples t_1, t_2,

$$\text{if } t_1[y_1] = t_2[y_1] \text{ then } \Pi_{y_2}(t_1[y_2], t_2[y_2]) \geq \Theta,$$

where Θ is a threshold value and $\Pi_{y_2}(t_1[y_2], t_2[y_2])$ is defined by

$$\Pi_{y_2}(t_1[y_2], t_2[y_2]) = \bigvee_{\langle u,v\rangle \in D_{y_2} \times D_{y_2}} \min((t_1[y_2])(u), u \sim_{y_2} v, (t_2[y_2])(v)).$$

Here, $\sim_{y_2}$ is a reflexive and symmetric fuzzy relation in D_{y_2}. One can easily see that if all fuzzy sets considered are normal, i.e. for every t and every y there exists $u \in D_y$ such that $(t[y])(u) = 1$, then Π_y is a reflexive and symmetric fuzzy relation on such fuzzy sets. Now, let $\mathcal{D}'$ be a ranked table with similarities such that the domain D'_y for $y \in Y$ in $\mathcal{D}'$ consists of all normal fuzzy sets in D_y, the similarity on D'_y is Π_y, and ranks are 1 for the tuples which are present in the original data table (and 0 for the others). One can easily verify the following assertion.

LEMMA 16
For * being globalization, $L = [0,1]$, and any $\rightarrow$, $\{y_1\} \Rightarrow \{y_2\}$ is true according to [29] iff $||\{y_1\} \Rightarrow \{y_2\}||_{\mathcal{D}'} \geq \Theta$, cf. (2).

That is, the semantics of the FDs from [29] can be defined in terms of the one defined by (2).

5 Conclusions

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 the 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. We showed that several of the approaches proposed in the literature are particular instances of our extension. Importantly, our article demonstrates that developing Codd's relational model with a clear connection to the framework of fuzzy logic in the narrow sense provides us with clarity, transparency and generality of the resulting model.

Two lines of research seem to be promising. From the point of view of database theory, further development of Codd's model of data (classic topics of the ordinary model such as relational algebra and calculus) based on fuzzy logic in the narrow sense is an obvious important goal. Note also that a thorough comparative study of further contributions, e.g. [12, 22, 24, 31, 33, 34], which are not covered in this article, is important in order to bring clarity to the topic. From the point of view of fuzzy logic, the situation is interesting as well. Namely, several interesting logical calculi have been studied for the ordinary Codd's model and their carrying over to a more general setting, e.g. to that of data tables with similarities, may results in interesting applied logical calculi.

Funding

Project MSM 6198959214 of the Czech Government.

References

[1] S. Abiteboul, R. Agrawal, P. Bernstein, M. Carey, S. Ceri, B. Croft, D. DeWitt, M. Franklin, H. G. Molina, D. Gawlick, J. Gray, L. Haas, A. Halevy, J. Hellerstein, Y. Ioannidis, M. Kersten, M. Pazzani, M. Lesk, D. Maier, J. Naughton, H. Schek, T. Sellis, A. Silberschatz, M. Stonebraker, R. Snodgrass, J. Ullman, G. Weikum, J. Widom, S. Zdonik. The Lowell database research self-assessment. *Comm. ACM* **48**, 111–118, 2005.

[2] W. W. Armstrong. Dependency structures in data base relationships. In *Information Processing: World Congress Proceedings*, pp. 580–583, North-Holland, Amsterdam, 1974.

[3] R. Belohlavek. *Fuzzy Relational Systems: Foundations and Principles*. Kluwer, Academic/Plenum Publishers, 2002.

[4] R. Belohlavek and V. Vychodil. Functional dependencies of data tables over domains with similarity relations. In *Proceedings of the 2nd Indian International Conference on Artificial Intelligence (IICAI'05)*, IICAI, pp. 2486–2504.

[5] R. Belohlavek and V. Vychodil. Attribute implications in a fuzzy setting. In *Proceedings of Formal Concept Analysis: 4th International Conference, (ICFCA'06)*, Vol. 3874 of *Lecture Notes in Artificial Intelligence*, R. Missaoui and J. Schmid eds, pp. 45–60, Springer, 2006.

[6] R. Belohlavek and V. Vychodil. Data tables with similarity relations: functional dependencies, complete rules and non-redundant bases. In *Proceedings of Database Systems for Advanced Applications: 11th International Conference, (DASFAA'06)*, Vol. 3882 of *Lecture Notes in Computer Science*, M. L. Lee, K. L. Tan and V. Wuwongse, eds, pp. 644–658, Springer, 2006.

[7] R. Belohlavek and V. Vychodil. Relational Model of Data over Domains with Similarities: An extension for similarity queries and knowledge extraction. In *The IEEE International Conference on Information Reuse and Integration*, pp. 207–213, IEEE, 2006.

[8] R. Belohlavek and V. Vychodil. Pavelka-style fuzzy logic for fuzzy attribute implications. In *JCIS*, *Proceedings of 9th Joint Conference on Information Sciences*, pp. 1152–1155, Atlantis Press, 2006.

[9] R. Belohlavek and V. Vychodil. Fuzzy attribute logic over complete residuated lattices. *Journal of Experimental and Theoretical Artificial Intelligence*, **18**, 471–480, 2006.

[10] B. Bhuniya and P. Niyogi. Lossless join property in fuzzy relational databases. *Data Knowledge Engineering*, **11**, 109–124, 1993.

[11] S. Ben Yahia, H. Ounalli and A. Jaoua. An extension of classical functional dependency: dynamic fuzzy functional dependency. *Information Sciences*, **119**, 219–234, 1999.

[12] P. Bosc, D. Dubois and H. Prade. Fuzzy functional dependencies. An overview and a critical discussion. In *FUZZ-IEEE'94*, pp. 325–330, IEEE , 1994.

[13] B. P. Buckles and F. E. Petry. A fuzzy representation of data for relational databases. *Fuzzy Sets and Systems*, **7**, 213–226, 1982.

[14] 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). *International Journal of Intelligent Systems*, **9**, 421–439, 1994.

[15] J. C. Cubero and M. A. Vila. A new definition of fuzzy functional dependency in fuzzy relational datatabses. *International Journal of Intelligent Systems*, **9**, 441–448, 1994.

[16] C. J. Date. *Database Relational Model: A Retrospective Review and Analysis.* Addison Wesley, 2000.

[17] G. Gerla. *Fuzzy Logic. Mathematical Tools for Approximate Reasoning.* Kluwer, 2001.

[18] S. Gottwald. *A Treatise on Many-Valued Logics.* Research Studies Press, 2001.

[19] P. Hájek. *Metamathematics of Fuzzy Logic.* Kluwer, 1998.

[20] P. Hájek. On very true. *Fuzzy Sets and Systems*, **124**, 329–333, 2001.

[21] I. F. Illyas, W. G. Aref and A. K. Elmagarmid. Supporting top-*k* join queries in relational databases. *The VLDB Journal*, **13**, 207–221, 2004.

[22] R. Intan and M. Mukaidono. Fuzzy conditional probability relations and their applications in fuzzy information systems. *Knowledge and Information Systems*, **6**, 345–365, 2004 .

[23] C. Li, K. C.-C. Chang, I. F. Ilyas and S. Song. RanSQL: Query algebra and Optimization for Relational top-k queries. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 131–142, ACM, 2005.

[24] W.-Y. Liu. Fuzzy data dependencies and implication of fuzzy data dependencies. *Fuzzy Sets and Systems*, **92**, 341–348, 1997.

[25] D. Maier. *The Theory of Relational Databases.* Computer Science Press, 1983.

[26] V. Novák, I. Perfilieva and J. Močkoř. *Mathematical Principles of Fuzzy Logic.* Kluwer, 1999.

[27] J. Pavelka. On fuzzy logic I, II, III. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **25**, 45–52, 119–134, 447–464, 1979.

[28] F. Petry. *Fuzzy Databases: Principles and Applications.* Kluwer Academic, 1996.

[29] H. Prade and C. Testemale. Generalizing database relational algebra for the treatment of incomplete or uncertain information and vague queries. *Information Sciences*, **34**, 115–143, 1984.

[30] K. V. S. V. N. Raju and A. K. Majumdar. Fuzzy functional dependencies and lossless join decomposition of fuzzy relational database systems. *ACM Transactions on Database Systems*, **13**, 129–166, 1988.

[31] S. Shenoi and A. Melton. Functional dependencies and normal forms in the fuzzy relational database model. *Information Sciences*, **100**, 1–28, 1992.

[32] G. Takeuti and S. Titani. Globalization of intuitionistic set theory. *Annals of Pure and Applied Logic*, **33**, 195–211, 1987.

[33] B. K. Tyagi, A. Sharfuddin, R. N. Dutta and D. K. Tayal. A complete axiomatization of fuzzy functional dependencies using fuzzy function. *Fuzzy Sets and Systems*, **151**, 363–379, 2005.

[34] S.-L. Wang, J.-S. Tsai and T.-P. Hong. Mining functional dependencies from fuzzy relational databases. In *ACM SAC 2000*, pp. 490–493, 2000.

Received 19 August 2008