ON CONSTRUCTION OF MRAP-SEQUENCES

R. BELOHLAVEK[1,2] and V. VYCHODIL[2]

[1] *Dept. Systems Science and Industrial Engineering, Binghamton University—SUNY; Binghamton, NY 13902, U. S. A.*
Email: rbelohla@binghamton.edu

[2] *Dept. Computer Science, Palacký University, Olomouc; Tomkova 40, CZ-779 00, Olomouc, Czech Republic*
Email: vilem.vychodil@upol.cz

Presented is an algorithm for constructing MRAP-sequences from sets of fuzzy attribute implications (FAIs). FAIs are formulas used to describe dependencies in object-attribute data with graded attributes as well as functional dependencies in ranked data tables over similarity relations. MRAP-sequences are particular normalized proofs from sets of FAIs. Using MRAP-sequences, one can discover dependencies following from other dependencies which is an important task in relational data analysis. We present theoretical foundations and the algorithm. All proofs are omited due to the limited scope of this paper.

Keywords: fuzzy attribute, fuzzy logic, if-then rule, normalized proof, residuated lattice

1. Introduction

In a series of papers we introduced fuzzy attribute logic which was originally motivated by problems in data analysis. Namely, we were interested in a tractable description and computation of non-redundant bases from given object-attribute data with fuzzy attributes, see [3] for a survey. Since this particular problem involves reasoning about entailment of dependencies (i.e., one needs to know whether certain dependency follows from other dependencies), we have developed fuzzy attribute logic, i.e. a logical calculus for reasoning with such dependencies. Fuzzy attribute logic (FAL) deals with rules, called fuzzy attribute implications, of the form "if A then B" where A and B are fuzzy sets of attributes, with the meaning: if an object has all the attributes of A then it has also all attributes of B. It is worth mentioning that fuzzy attribute implications also play an important role in a similarity-based extension of Codd's relational database model proposed in [4]. In this case, FAIs play the role of functional dependencies in ranked data tables with similarity relations. The two ways of interpreting fuzzy attribute implications, i.e. in data tables with fuzzy attributes and in ranked data tables with similarity relations, are closely connected. Namely, semantic entailment coincides for both of them.

In [2,6] we showed several axiomatizations of FAL and proved their completeness over arbitrary complete residuated lattices with hedges taken as structures of truth degrees. In [7] we have paid attention to one of our deduction rules and, as a consequence of its properties, we obtained several ways to normalize proofs in FAL. Using the methods in [7], one can take a proof and convert it to a (normalized) proof in certain form. Hence, in order to obtain a normalized proof of a FAI, one already has to have at least one proof of that FAI. Needless to say, finding proofs can be a tedious procedure, i.e., it is desirable to have an algorithm—if there is one—which can find a (normalized) proof automatically or answers that a proof cannot be found if the FAI is not provable. In this paper we show that this is indeed possible due to properties of least models studied in [5]. Moreover, our algorithm generates proofs which are in a normal form, so-called MRAP-sequences. These sequences generalize RAP-sequences well known from database systems [11]. The presented algorithm can be used to determine whether a FAI follows from a given collection of FAIs which can be used, e.g., in algorithms for computing non-redundant bases.

Our paper is organized as follows. Section 2 contains notions of FAL necessary to present the main result, which is the algorithm in Section 3 and the proof of its correctness.

Preliminaries We use complete residuated lattices with truth-stressing hedges (shortly, hedges) as structures of truth degrees. In the following we use well-known properties of residuated lattices with hedges (denoted $\mathbf{L}$) and fuzzy structures which can be found in [1,9,10].

2. Fuzzy Attribute Logic (FAL)

In this section we recall basic notions of fuzzy attribute logic (FAL). More details can be found in survey papers [3,4] and in the references therein.

2.1. *Formulas of FAL and their semantics*

Let Y be a finite set of attributes. A *fuzzy attribute implication* (*over* Y) is an expression $A \Rightarrow B$, where $A, B \in \mathbf{L}^Y$ (A and B are fuzzy sets of attributes). We consider validity (truth) of FAIs as follows: For an **L**-set $M \in \mathbf{L}^Y$ of attributes, we define a *degree* $||A \Rightarrow B||_M \in L$ *to which* $A \Rightarrow B$ *is true in* M by $||A \Rightarrow B||_M = S(A, M)^* \rightarrow S(B, M)$, where $S(\cdots)$ denote subsethood degrees [1]. The degree $||A \Rightarrow B||_M$ can be understood as follows: if M (semantic component) represents presence of attributes of some object, i.e. $M(y)$ is truth degree to which "the object has the attribute $y \in Y$", then $||A \Rightarrow B||_M$ is the truth degree to which "if the object has all attributes from A, then it has all attributes from B", which corresponds to the desired interpretation of $A \Rightarrow B$.

Let T be a set of fuzzy attribute implications. $M \in \mathbf{L}^Y$ is called a *model of* T if $||A \Rightarrow B||_M = 1$ for each $A \Rightarrow B \in T$. The set of all models of T is denoted by $\mathrm{Mod}(T)$. A *degree* $||A \Rightarrow B||_T \in L$ *to which* $A \Rightarrow B$ *semantically follows from* T is defined by $||A \Rightarrow B||_T = \bigwedge_{M \in \mathrm{Mod}(T)} ||A \Rightarrow B||_M$. The degree $||A \Rightarrow B||_T$ of semantic entailment is defined as a degree to which "$A \Rightarrow B$ is true in each model of T". If $||A \Rightarrow B||_T = 1$ we say that $A \Rightarrow B$ *is fully entailed by* T.An important property of the semantic entailment is that it can be characterized by least models [5]. In a more detail, each system $\mathrm{Mod}(T)$ of models is closed under arbitrary intersections, i.e., for each $M \in \mathbf{L}^Y$ we can consider the least model $cl_T(M)$ of T (least with respect to "$\subseteq$") containing M. From [5] we have the following observation:

Theorem 2.1 (see [5]). *Let T be a set of FAIs. Then, for each $A \Rightarrow B$, we have $||A \Rightarrow B||_T = S(B, cl_T(A))$.* □

2.2. *Deductive systems of FAL*

In this section we focus on syntactic entailment (provability) from sets of FAIs. In what follows we consider *deduction rules* of the form "from $\varphi_1, \ldots, \varphi_n$ infer φ", where n is a nonnegative integer and φ, φ_i $(i \in I)$ are (schemas for) FAIs. Such deduction rules are to be understood as usual: having rule "from $\varphi_1, \ldots, \varphi_n$ infer φ" and FAIs which are of the form of the FAIs in the input part (the part preceding "infer") of the rule, the rule allows us to infer (in one step) the corresponding FAI in the output part (the part following "infer") of the rule. Each nullary rule, i.e. rule where $n = 0$, is considered as an axiom the output part of which can be inferred in one step.

Deductive system of FAL [3] uses the following 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 $c \in L$. The rules are inspired by Armstrong axioms well known from database systems [11]. Namely, (Ax) and (Cut) are counterparts of the ordinary Armstrong rules with sets of attributes replaced by fuzzy sets. (Mul) is a new rule called *multiplication* and does not have a nontrivial classical counterpart. Recall that $c^* \otimes A$ and $c^* \otimes B$ which appear in (Mul) represent c^*-multiples of A and B, respectively.

A fuzzy attribute implication $A \Rightarrow B$ is called *provable from a set T of FAIs using a set $\mathcal{R}$ of deduction rules* if there is a sequence $\varphi_1, \ldots, \varphi_n$ of fuzzy attribute implications such that φ_n is $A \Rightarrow B$ and for each φ_i we either have $\varphi_i \in T$ or φ_i is inferred (in one step) from some of the preceding formulas using some deduction rule from $\mathcal{R}$ (i.e., $\mathcal{R}$ contains a rule "from $\psi_1, \ldots, \psi_k$ infer φ_i" where each of $\psi_1, \ldots, \psi_k$ is among $\varphi_1, \ldots, \varphi_{i-1}$). The sequence $\varphi_1, \ldots, \varphi_n$ is then called a *proof of φ from T using $\mathcal{R}$*. If $\mathcal{R}$ consists of (Ax)–(Mul), we say just "$A \Rightarrow B$ is provable from T" instead of "$A \Rightarrow B$ is provable from T using ..." and denote this fact by $T \vdash A \Rightarrow B$. A basic relationship between semantic entailment and provability is given by the following theorem (see [2,3]):

Theorem 2.2 (completeness). *Let T be a set of FAIs. Then, for each $A \Rightarrow B$, we have $T \vdash A \Rightarrow B$ iff $||A \Rightarrow B||_T = 1$.* □

The first results on completeness of FAL were published in [2]. Note that the original results were restricted to finite structures of truth degrees and later we showed that axiomatization over infinite structures is also possible, see [6]. Theorem 2.2 characterizes FAIs which are fully entailed by T, i.e. FAIs which follow from T to degree 1. In addition to that, we can prove that FAI is complete in a graded style. That is, we have a notion of a provability degree $|A \Rightarrow B|_T \in L$ of $A \Rightarrow B$ from T such that $|A \Rightarrow B|_T = ||A \Rightarrow B||_T$ is true for any $A \Rightarrow B$ and T. It is also possible to consider semantic entailment from fuzzy sets of FAIs instead of the ordinary sets. For details we refer to [2,3].

From the viewpoint of applications, it is important to have a clear procedure for finding proofs of FAIs from collection of other FAIs. If available, such a procedure could be used for a variety of problems. For instance, it could be used in algorithms for generating non-redundant bases of data tables [3,4]. A step towards the desired procedure has been made in [7] where we have introduced several types of normalized proofs. An important type of normalized proofs is based on deductive rules of *reflexivity*, *accumulation*, and *projectivity* which are inspired by analogous rules from database systems [11] and which replace the rules (Ax) and (Cut). In our setting, we define such rules as follows:

(Ref) infer $A \Rightarrow A$,

(Acc) from $A \Rightarrow B \cup C$ and $C \Rightarrow D \cup E$ infer $A \Rightarrow B \cup C \cup D$,

(Pro) from $A \Rightarrow B \cup C$ infer $A \Rightarrow B$,

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

An *MRAP-sequence* for $A \Rightarrow B$ from T is a proof of $A \Rightarrow B$ from T using (Mul), (Ref), (Acc), (Pro), such that

1. the proof starts with FAIs from T,
2. continues with FAIs which result by application of (Mul) to formulas from 1.,
3. continues with $A \Rightarrow A$, i.e. with an instance of (Ref),
4. continues with formulas which result by application of (Acc) to formulas from 1., 2., and 3.,
5. ends with application of (Pro) which results in $A \Rightarrow B$, the last member of the proof.

Observe that (Ref) and (Pro) are used in each MRAP-sequence just once while (Mul) and (Add) can be used several times. All applications of (Mul) appear in a single block (see 2.), the same applies for (Add), see 4. In [7] we have proved the following assertion extending Theorem 2.2:

Theorem 2.3 (MRAP-sequence theorem). *Let T be a set of FAIs. Then the following are equivalent:*

1. $||A \Rightarrow B||_T = 1$,
2. $A \Rightarrow B$ *is provable from* T *using* (Mul), (Ax), *and* (Cut),
3. *there is an MRAP-sequence for* $A \Rightarrow B$ *from* T. □

3. Construction of MRAP-sequences

In this section we focus on computing MRAP-sequences from given T. First, we introduce sequences of fuzzy sets of attributes which will be further used to obtain MRAP-sequences. For brevity, we consider the following notation: For fuzzy sets $A, B \in \mathbf{L}^Y$ and a FAI $C \Rightarrow D$ we denote by $A \subset_{C \Rightarrow D} B$ the fact that

(i) $A \subset B$, and

(ii) $B = A \cup (S(C, A)^* \otimes D)$.

The relationship between A and B prescribed by (ii) can be explained as follows: an attribute y belongs to B to a degree to which y belongs to A, or C is (very) contained in A and y belongs to D. The next assertion says that for each set of FAIs and an initial fuzzy set of attributes we can form particular chain of fuzzy sets of attributes which, together with certain FAIs, satisfy (i) and (ii).

Theorem 3.1. *Let* $\mathbf{L}$ *be finite,* T *be a set of FAIs,* $A \in \mathbf{L}^Y$ *be a fuzzy set of attributes,* $cl_T(A)$ *be the least model of* T *containing* A*. Then there are fuzzy sets* $A_0, A_1, \ldots, A_n \in \mathbf{L}^Y$ *and FAIs* $\varphi_0, \ldots, \varphi_{n-1} \in T$ *such that*

1. $A_0 = A$, $A_n = cl_T(A)$, *and*
2. *for each* $i = 1, \ldots, n-1$: $A_i \subset_{\varphi_i} A_{i+1}$. □

The previous assertion will be used to justify the following algorithm for computing MRAP-sequences from T.

Algorithm 3.1 (Compute MRAP-sequence).

Input: set T of FAIs and $A \Rightarrow B$
Output: "No" or "Yes" + MRAP-seq. for $A \Rightarrow B$ from T

set i ***to*** 0; ***set*** A_0 ***to*** A
while there is $C_i \Rightarrow D_i \in T$ ***such that*** $S(C_i, A_i)^* \nleq S(D_i, A_i)$:
  set φ_i ***to*** $C_i \Rightarrow D_i$
  set A_{i+1} ***to*** $A_i \cup (S(C_i, A_i)^* \otimes D_i)$
  increase i ***by*** 1
set n ***to*** i
if $B \nsubseteq A_n$:
  return "No"
return "Yes" + the following sequence:

$$
\begin{aligned}
&C_0 \Rightarrow D_0, C_1 \Rightarrow D_1, \ldots, C_{n-1} \Rightarrow D_{n-1},\\
&S(C_0, A_0)^* \otimes C_0 \Rightarrow S(C_0, A_0)^* \otimes D_0,\\
&S(C_1, A_1)^* \otimes C_1 \Rightarrow S(C_1, A_1)^* \otimes D_1,\\
&\quad\vdots\\
&S(C_{n-1}, A_{n-1})^* \otimes C_{n-1} \Rightarrow S(C_{n-1}, A_{n-1})^* \otimes D_{n-1},\\
&A \Rightarrow A_0, A \Rightarrow A_1, \ldots, A \Rightarrow A_{n-1}, A \Rightarrow A_n,\\
&A \Rightarrow B.
\end{aligned}
$$

Theorem 3.2 (correctness of Algorithm 3.1).
For each T *and* $A \Rightarrow B$*, Algorithm 3.1 returns (i) "Yes" if* $||A \Rightarrow B||_T = 1$ *and the resulting sequence is an MRAP-sequence for* $A \Rightarrow B$ *from* T*; (ii) returns "No" otherwise.* □

4. Remarks and Conclusions

In this paper we have introduced an algorithm for computing MRAP-sequences. The algorithm can be used to determine whether an if-then dependency follows from other dependencies. The dependencies we use are fuzzy attribute implications (FAIs). If we choose a two-value Boolean algebra as a structure of truth degrees, FAIs become the ordinary attribute implications [8] with their two-valued interpretation. Therefore, our algorithm pertains to the classical attribute implications as well. The algorithm can also be used to determine functional dependencies in database tables in sense of Codd's relational database model [11] and its similarity-based extension presented in [4]. It can be shown that the asymptotic complexity of our algorithm is in $O(nk)$, where n is the size of T and k is the number of fixed points of the hedge *. Our future research will focus on further types of data dependencies in graded setting and related issues.

Acknowledgement 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.

References

1. R. Belohlavek: *Fuzzy Relational Systems: Foundations and Principles.* Kluwer, Academic/Plenum Publishers, New York, 2002.
2. R. Belohlavek, V. Vychodil: Fuzzy attribute logic: syntactic entailment and completeness. In: *Proc. JCIS 2005*, pp. 78–81, ISBN 0–9707890–3–3.
3. R. Belohlavek, V. Vychodil: Attribute implications in a fuzzy setting. In: Missaoui R., Schmid J. (Eds.): *ICFCA 2006*, LNAI **3874**, pp. 45–60, 2006.
4. R. Belohlavek, V. Vychodil: Data tables with similarity relations: functional dependencies, complete rules and non-redundant bases. In: Lee, Tan, and Wuwongse (Eds.): *DASFAA 2006*, LNCS **3882**, pp. 644–658, 2006.
5. R. Belohlavek, V. Vychodil: Properties of models of fuzzy attribute implications. In: *Proc. SCIS & ISIS 2006*, pp. 291–296, ISSN 1880–3741.
6. R. Belohlavek, V. Vychodil: Fuzzy attribute logic over complete residuated lattices. *J. Exp. Theor. Artif. Intelligence* **18**(4)(2006), pp. 471–480, 2006.
7. R. Belohlavek, V. Vychodil: On proofs and rule of multiplication in fuzzy attribute logic (to appear in *Proc. IFSA 2007*).
8. B. Ganter, R. Wille: *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin, 1999.
9. P. Hájek: *Metamathematics of Fuzzy Logic.* Kluwer, Dordrecht, 1998.
10. P. Hájek: On very true. *Fuzzy Sets and Systems* **124**(2001), 329–333.
11. D. Maier: *The Theory of Relational Databases.* Computer Science Press, Rockville, 1983.