Computing non-redundant bases of if-then rules from data tables with graded attributes

Radim Bělohlávek and Vilém Vychodil
Dept. Computer Science, Palacký University, Tomkova 40, CZ-779 00, Olomouc, Czech Republic
Email: {radim.belohlavek, vilem.vychodil}@upol.cz

***Abstract*— We present a method for computation of non-redundant bases of attribute implications from data tables with fuzzy attributes. Attribute implications are formulas describing particular dependencies of attributes in data. A non-redundant basis is a minimal set of attribute implications such that each attribute implication which is true in a given data (semantically) follows from the basis. Our bases are uniquely given by so-called systems of pseudo-intents. Pseudo-intents are particular granules in data tables. We reduce the problem of computing systems of pseudo-intents to the problem of computing maximal independent sets in certain graphs. We present theoretical foundations, the algorithm, and demonstrating examples.**

I. Introduction

Fuzzy attribute implications are formulas of a form $A \Rightarrow B$ which describe particular dependencies in data tables. Among all the implications which are true in a given table, there is a lot of them which are trivial and a lot of them which can be removed since they are entailed by the others. Therefore, it is desirable to look for methods for obtaining non-redundant bases, i.e. minimal sets of attribute implications such that all implications which are true in a given data table semantically follow from the basis. In case of data tables with binary attributes (tables containing 0's and 1's), an algorithm for computation of non-redundant bases is known, see [11], [13]. For data tables with fuzzy attributes (tables containing degrees, e.g. reals from $[0,1]$), non-redundant bases of implications determined by so-called systems of pseudo-intents are described in [6], [9]. For a particular case when a hedge (a unary logical function used in the definition of validity of attribute implications) is a so-called globalization on a finite scale of truth degrees, there is a unique system of pseudo-intents and an algorithm for the computation of the corresponding non-redundant basis was presented in [6]. For general hedges, however, there might be several systems of pseudo-intents of a given table with fuzzy attributes. Up to now, it was not known how to compute the systems of pseudo-intents and the corresponding non-redundant bases for other hedges than globalization.

In the present paper, we show an algorithm to compute systems of pseudo-intents and the corresponding non-redundant bases for a general case (i.e., for any hedge). The pseudo-intents can be thought of as particular granules in data which can be used to determine a minimal fully informative set of attribute implications. The algorithm is based on a theorem according to which the systems of pseudo-intents correspond to maximal independent sets in a particular graph. We present theoretical foundations, the algorithm, its relationship to previously published algorithm [6] which works for a hedge being globalization, and demonstrating examples. Preliminary results on this topic were presented in [10].

II. Preliminaries

Fuzzy logic and fuzzy set theory are formal frameworks for a manipulation of a particular form of imperfection called fuzziness (vagueness). Contrary to classical logic, fuzzy logic uses a scale L of truth degrees, a most common choice being $L = [0,1]$ (real unit interval) or some subchain of $[0,1]$. This enables us to consider intermediate truth degrees of propositions, e.g. "object x has attribute y" has a truth degree 0.8 indicating that the proposition is almost true. In addition to a set L of truth degrees, one has to pick an appropriate collection of logical connectives (implication, conjunction, $\dots$). A general choice of a set of truth degrees plus logical connectives is represented by so-called complete residuated lattices (equipped possibly with additional operations). The rest of this section presents an introduction to fuzzy logic notions we need in the sequel. Details can be found e.g. in [4], [12], [14], a good introduction to fuzzy logic and fuzzy sets is presented in [17].

A complete residuated lattice with a truth-stressing hedge (shortly, a hedge) [14], [15] 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 $a \leq b \rightarrow c$ for each $a, b, c \in L$; hedge * satisfies

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

for each $a, b \in L$, $a_i \in L$ ($i \in I$). Elements a of L are called truth degrees. $\otimes$ and $\rightarrow$ are (truth functions of) "fuzzy conjunction" and "fuzzy implication". Hedge * is a (truth function of) logical connective "very true", see [14], [15]. Properties (2)–(4) have natural interpretations, e.g. (2) can be read: "if a is very true, then a is true", (3) can be read: "if $a \rightarrow b$ is very true and if a is very true, then b is very true", etc.

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

$$\text{Łukasiewicz:} \quad \begin{aligned} a \otimes b &= \max(a + b - 1, 0), \\ a \rightarrow b &= \min(1 - a + b, 1), \end{aligned} \tag{5}$$

$$\text{Gödel:} \quad \begin{aligned} a \otimes b &= \min(a, b), \\ a \rightarrow b &= \begin{cases} 1 & \text{if } a \leq b, \\ b & \text{otherwise,} \end{cases} \end{aligned} \tag{6}$$

$$\text{Goguen (product):} \quad \begin{aligned} a \otimes b &= a \cdot b, \\ a \rightarrow b &= \begin{cases} 1 & \text{if } a \leq b, \\ \frac{b}{a} & \text{otherwise.} \end{cases} \end{aligned} \tag{7}$$

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

Gödel chain which consists of L and restrictions of Gödel operations on $[0,1]$ to L.

Two boundary cases of (truth-stressing) hedges are (i) identity, i.e. $a^* = a$ ($a \in L$); (ii) globalization [19]:

$$a^* = \begin{cases} 1 & \text{if } a = 1, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

A special case of a complete residuated lattice with hedge is the two-element Boolean algebra $\langle \{0,1\}, \wedge, \vee, \otimes, \rightarrow, {}^*, 0, 1 \rangle$, denoted by **2**, which is the structure of truth degrees of the classical logic. That is, the operations $\wedge, \vee, \otimes, \rightarrow$ of **2** are the truth functions (interpretations) of the corresponding logical connectives of the classical logic and $0^* = 0$, $1^* = 1$.

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

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

$$S(A,B) = \textstyle\bigwedge_{u \in U} \big(A(u) \rightarrow B(u)\big), \tag{9}$$

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

In the following we use well-known properties of residuated lattices and fuzzy structures which can be found in monographs [4], [14]. Throughout the rest of the paper, $\mathbf{L}$ denotes an arbitrary complete residuated lattice with a hedge.

III. Fuzzy attribute logic

A. Fuzzy attribute implications and their validity

We first introduce attribute implications. Suppose Y is a finite set of attributes. A (*fuzzy*) *attribute implication* (*over attributes* Y) is an expression $A \Rightarrow B$, where $A, B \in \mathbf{L}^Y$ (A and B are fuzzy sets of attributes). The intended meaning of $A \Rightarrow B$ is: "if it is (very) true that an object has all attributes from A, then it has also all attributes from B" with the logical connectives being given by $\mathbf{L}$. A fuzzy attribute implication does not have any kind of "validity" on its own—it is a syntactic notion. In order to consider validity, we must introduce an interpretation of fuzzy attribute implications. Fuzzy attribute implications are meant to be interpreted in data tables with fuzzy attributes [6], [7], [9]. A data table with fuzzy attributes can be seen as a triplet $\mathcal{T} = \langle X, Y, I \rangle$ where X is a set of objects, Y is a finite set of attributes (the same as above in the definition of a fuzzy attribute implication), and $I \in \mathbf{L}^{X \times Y}$ is a binary $\mathbf{L}$-relation between X and Y assigning to each object $x \in X$ and each attribute $y \in Y$ a degree $I(x,y)$ to which x has y. $\mathcal{T} = \langle X, Y, I \rangle$ can be thought as a table with rows and columns corresponding to objects $x \in X$ and attributes $y \in Y$, respectively, and table entries containing degrees $I(x,y)$. A row of a table $\mathcal{T} = \langle X, Y, I \rangle$ corresponding to an object $x \in X$ can be seen as a fuzzy set I_x of attributes to which an attribute $y \in Y$ belongs to a degree $I_x(y) = I(x,y)$. For fuzzy set $M \in \mathbf{L}^Y$ of attributes, we define a *degree* $||A \Rightarrow B||_M \in L$ *to which* $A \Rightarrow B$ *is true in* M by

$$||A \Rightarrow B||_M = S(A,M)^* \rightarrow S(B,M). \tag{10}$$

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

$$||A \Rightarrow B||_{\mathcal{M}} = \textstyle\bigwedge_{M \in \mathcal{M}} ||A \Rightarrow B||_M. \tag{11}$$

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

B. Completeness and non-redundant bases

Let $\mathcal{T} = \langle X, Y, I \rangle$ be a data table with fuzzy attributes and 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 $||A \Rightarrow B||_T$ to which $A \Rightarrow B$ *semantically follows* from T is defined by

$$||A \Rightarrow B||_T = ||A \Rightarrow B||_{\mathrm{Mod}(T)}. \tag{12}$$

T is called *complete* (in $\mathcal{T} = \langle X, Y, I \rangle$) if $||A \Rightarrow B||_T = ||A \Rightarrow B||_{\langle X,Y,I \rangle}$, i.e. if a degree to which T entails $A \Rightarrow B$ coincides with a degree to which $A \Rightarrow B$ is true in $\langle X, Y, I \rangle$, for each $A \Rightarrow B$. If T is complete and no proper subset of T is complete, then T is called a *non-redundant basis* (of $\mathcal{T} = \langle X, Y, I \rangle$). Note that both the notions of a complete set and a non-redundant basis refer to a given data table with fuzzy attributes.

C. Fuzzy concept lattices with hedges

Given a data table with fuzzy attributes $\mathcal{T} = \langle X, Y, I \rangle$, for $A \in \mathbf{L}^X$, $B \in \mathbf{L}^Y$ (i.e. A is a fuzzy set of objects, B is a fuzzy set of attributes), we define fuzzy sets $A^\uparrow \in \mathbf{L}^Y$ (fuzzy set of attributes), $B^\downarrow \in \mathbf{L}^X$ (fuzzy set of objects) by

$$A^\uparrow(y) = \textstyle\bigwedge_{x \in X} \big(A(x)^* \rightarrow I(x,y)\big), \tag{13}$$

$$B^\downarrow(x) = \textstyle\bigwedge_{y \in Y} \big(B(y) \rightarrow I(x,y)\big). \tag{14}$$

We put

$$\mathcal{B}(X^*, Y, I) = \{\langle A, B \rangle \in \mathbf{L}^X \times \mathbf{L}^Y \mid A^\uparrow = B,\ B^\downarrow = A\}$$

and define for $\langle A_1, B_1 \rangle, \langle A_2, B_2 \rangle \in \mathcal{B}(X^*, Y, I)$ a binary relation $\leq$ by $\langle A_1, B_1 \rangle \leq \langle A_2, B_2 \rangle$ iff $A_1 \subseteq A_2$ (or, iff $B_2 \subseteq B_1$; both ways are equivalent). Operators ${}^\downarrow, {}^\uparrow$ form so-called Galois connection with hedge, see [8]. The structure $\langle \mathcal{B}(X^*, Y, I), \leq \rangle$ is called a *fuzzy concept lattice* induced by $\mathcal{T} = \langle X, Y, I \rangle$. The elements $\langle A, B \rangle$ of $\mathcal{B}(X^*, Y, I)$ are naturally interpreted as concepts (clusters) hidden in the input data represented by I. Namely, $A^\uparrow = B$ and $B^\downarrow = A$ say that B is the collection of all attributes shared by all objects from A, and A is the collection of all objects sharing all attributes from B. Note that these conditions represent exactly the definition of a concept as developed in the so-called Port-Royal logic; A and B are called the *extent* and the *intent* of the concept $\langle A, B\rangle$, respectively, and represent the collection of all objects and all attributes covered by the particular concept. Furthermore, $\leq$ models the natural subconcept-superconcept hierarchy—concept $\langle A_1, B_1\rangle$ is a subconcept of $\langle A_2, B_2\rangle$ iff each object from A_1 belongs to A_2 (dually for attributes).

D. Systems of pseudo-intents

Defnition 1: Given a data table with fuzzy attributes $\mathcal{T} = \langle X, Y, I\rangle$, a system of fuzzy sets of attributes $\mathcal{P} \subseteq \mathbf{L}^Y$ is called a *system of pseudo-intents of* $\mathcal{T}$ if for each $P \in \mathbf{L}^Y$ we have:

$$P \in \mathcal{P} \quad \text{iff} \quad P \neq P^{\downarrow\uparrow} \text{ and for each } Q \in \mathcal{P} \text{ such that } Q \neq P \text{ we have } ||Q \Rightarrow Q^{\downarrow\uparrow}||_P = 1. \tag{15}$$

The following property of pseudo-intents was proved in [9]:

Theorem 1: Let $\mathcal{T} = \langle X, Y, I\rangle$ be a data table with fuzzy attributes. If $\mathcal{P}$ is a system of pseudo-intents of $\mathcal{T}$ then

$$T = \{P \Rightarrow P^{\downarrow\uparrow} \mid P \in \mathcal{P}\} \tag{16}$$

is a non-redundant basis of $\mathcal{T}$. ■

If * is globalization, there is a unique system of pseudo-intents and an efficient algorithm exists for its computation [6]. In what follows we will be concerned with the case of a general *.

IV. Computing non-redundant bases

For any data table with fuzzy attributes $\mathcal{T} = \langle X, Y, I\rangle$ we define a set V of $\mathbf{L}$-sets of attributes by

$$V = \{P \in \mathbf{L}^Y \mid P \neq P^{\downarrow\uparrow}\}. \tag{17}$$

Now, in case of empty V, there is exactly one system of pseudo-intents:

Lemma 1: If V defined by (17) is empty then $\mathcal{P} = \emptyset$ is the only system of pseudo-intents of $\mathcal{T} = \langle X, Y, I\rangle$.

Proof: If $V = \emptyset$ then $\mathcal{P} = \emptyset$ trivially satisfies (15). The rest of the claim follows from the fact that each system of pseudo-intents is a subset of V. ■

Note that the case of empty V is not interesting from the user's point of view: $V = \emptyset$ means that any $\mathbf{L}$-set of attributes is an intent of some cluster extracted from $\mathcal{T} = \langle X, Y, I\rangle$. In other words, any $\mathbf{L}$-set of attributes forms a cluster in $\mathcal{T}$. Thus, in the sequel we will be interested in systems of pseudo-intents for $V \neq \emptyset$. For nonempty V we define a binary relation E on V by

$$E = \{\langle P, Q\rangle \in V \mid P \neq Q \text{ and } ||Q \Rightarrow Q^{\downarrow\uparrow}||_P \neq 1\}. \tag{18}$$

In case of nonempty V, $\mathbf{G} = \langle V, E \cup E^{-1}\rangle$ is a graph. For any $Q \in V$ and $\mathcal{P} \subseteq V$ define the following subsets of V:

$$\mathrm{Pred}(Q) = \{P \in V \mid \langle P, Q\rangle \in E\},$$
$$\mathrm{Pred}(\mathcal{P}) = \textstyle\bigcup_{Q \in \mathcal{P}} \mathrm{Pred}(Q).$$

Described verbally, $\mathrm{Pred}(Q)$ is a set of all elements from V which are predecessors of Q (in E). $\mathrm{Pred}(\mathcal{P})$ is a set of all predecessors of any $Q \in \mathcal{P}$.

Lemma 2: Let $\emptyset \neq \mathcal{P} \subseteq \mathbf{L}^Y$. If $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$ then $\mathcal{P}$ is a maximal independent set in $\mathbf{G}$.

Proof: $\mathcal{P}$ is independent: Take $P, Q \in \mathcal{P}$. If $\langle P, Q\rangle \in E$, we would obtain $P \in \mathrm{Pred}(Q)$, i.e. $P \in \mathrm{Pred}(\mathcal{P}) = V - \mathcal{P}$ which contradicts the fact that $P \in \mathcal{P}$.

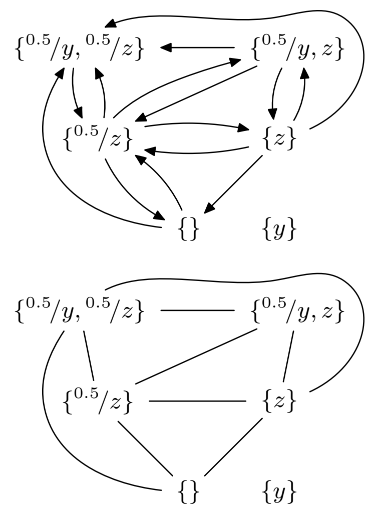


Fig. 1. Graphs induced by data table from Example 1

$\mathcal{P}$ is maximal: Take $Q \in V$ such that $Q \notin \mathcal{P}$. Since $Q \in V - \mathcal{P}$ and $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$, there is $P \in \mathcal{P}$ such that $Q \in \mathrm{Pred}(P)$, i.e. $\langle Q, P\rangle \in E$. Hence, Q cannot be added to $\mathcal{P}$ without loosing the independence. ■

Theorem 2: Let $\mathcal{P} \subseteq \mathbf{L}^Y$. $\mathcal{P}$ is a system of pseudo-intents iff $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$.

Proof: "$\Rightarrow$": Let $\mathcal{P}$ be a system of pseudo-intents. We prove both inclusions of $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$. Observe that for $P, Q \in \mathcal{P}$ we have $\langle P, Q\rangle \notin E$ because either $P = Q$, or $P \neq Q$ and thus $||Q \Rightarrow Q^{\downarrow\uparrow}||_P = 1$ by (15). Therefore, $P \notin \mathrm{Pred}(Q)$ for each $P, Q \in \mathcal{P}$, i.e. $P \notin \mathrm{Pred}(\mathcal{P})$ for each $P \in \mathcal{P}$. Thus, for $P \in \mathrm{Pred}(\mathcal{P})$ we have $P \notin \mathcal{P}$, which gives $P \in V - \mathcal{P}$. We have shown $\mathrm{Pred}(\mathcal{P}) \subseteq V - \mathcal{P}$. Conversely, take $P \in V - \mathcal{P}$, i.e. $P \neq P^{\downarrow\uparrow}$ and $P \notin \mathcal{P}$. Since $\mathcal{P}$ is a system of pseudo-intents, there is $Q \in \mathcal{P}$ such that $P \neq Q$ and $||Q \Rightarrow Q^{\downarrow\uparrow}||_P \neq 1$, i.e. $\langle P, Q\rangle \in E$. That is, $P \in \mathrm{Pred}(Q)$ where $Q \in \mathcal{P}$, yielding $P \in \mathrm{Pred}(\mathcal{P})$. Thus, we have $V - \mathcal{P} \subseteq \mathrm{Pred}(\mathcal{P})$. Altogether, $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$.

"$\Leftarrow$": Suppose we have $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$. Take any $P \in \mathbf{L}^Y$. We have $P \in \mathcal{P}$ iff $P \in V$ and $P \notin V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$ which is true iff $P \neq P^{\downarrow\uparrow}$ and $\langle P, Q\rangle \notin E$ for all $Q \in \mathcal{P}$. The latter is true iff $P \neq P^{\downarrow\uparrow}$ and, for each $Q \in \mathcal{P}$, if $P \neq Q$ then $||Q \Rightarrow Q^{\downarrow\uparrow}||_P = 1$. Since $P \in \mathbf{L}^Y$ was taken arbitrarily, (15) gives that $\mathcal{P}$ is a system of pseudo-intents. ■

Lemma 2 and Theorem 2 yield

Corollary 1: $\mathcal{P} \neq \emptyset$ is a system of pseudo-intents iff $\mathcal{P}$ is a maximal independent set in $\mathbf{G}$ such that $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$. ■

Remark 1: Corollary 1 provides a way to compute systems of pseudo-intents. It suffices to find all maximal independent sets in $\mathbf{G}$ and check which of them satisfy additional condition $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$ (this property can be checked during the generating of maximal independent sets). The procedure is illustrated by the following example. For more examples see Section V.

Example 1: Let $\mathbf{L}$ be a three-element Łukasiewicz chain with $L = \{0, 0.5, 1\}$, and * being identity. Consider a data table with fuzzy attributes $\mathcal{T} = \langle X, Y, I\rangle$ where $X = \{x\}$, $Y = \{y, z\}$, $I(x, y) = 0.5$, and $I(x, z) = 0$. For $\mathcal{T}$, V defined by (17) is the following

$$V = \{\{\}, \{^{0.5}/z\}, \{z\}, \{^{0.5}/y, {}^{0.5}/z\}, \{^{0.5}/y, z\}, \{y\}\}.$$

procedure: UPDATE
input: sequence $\mathcal{S} = P_0, \ldots, P_k$
output: sequence $\mathcal{S}'$

vacate $\mathcal{S}'$, *set* i *to* 1
while $i \leq k$:
  if $\langle P_i, P_0 \rangle \notin E$:
    append P_i *to* $\mathcal{S}'$
  set i *to* $i+1$

algorithm: GBASIS
input: $E \subseteq V \times V$, sequence $\mathcal{S}$
output: subset $\mathcal{P} \subseteq V$

set $\mathcal{P}$ *to* $\emptyset$
while $\mathcal{S}$ *is not empty*:
  add the first element of $\mathcal{S}$ *to* $\mathcal{P}$
  set $\mathcal{S}$ *to* $\mathrm{UPDATE}(\mathcal{S})$

Fig. 2. Procedure UPDATE and algorithm GBASIS

The corresponding binary relation E defined by (18) is depicted in Fig. 1 (top); graph $\mathbf{G} = \langle V, E \cup E^{-1} \rangle$ is depicted in Fig. 1 (bottom). $\mathbf{G}$ contains four maximal independent sets:

$$\mathcal{P}_1 = \{\{\}, \{{}^{0.5}/y, z\}, \{y\}\}, \qquad \mathcal{P}_3 = \{\{z\}, \{y\}\},$$
$$\mathcal{P}_2 = \{\{{}^{0.5}/z\}, \{y\}\}, \qquad \mathcal{P}_4 = \{\{{}^{0.5}/y, {}^{0.5}/z\}, \{y\}\}.$$

Observe that $\mathcal{P}_1$ and $\mathcal{P}_3$ do not satisfy $V - \mathcal{P}_i = \mathrm{Pred}(\mathcal{P}_i)$ ($i \in \{1,3\}$) because $\{{}^{0.5}/y, {}^{0.5}/z\} \notin \mathrm{Pred}(\mathcal{P}_i)$ ($i \in \{1,3\}$), and $\{\} \notin \mathrm{Pred}(\mathcal{P}_3)$. Thus, $\mathcal{P}_1$ and $\mathcal{P}_3$ are not systems of pseudo-intents of $\mathcal{T}$ due to Theorem 2. On the other hand, $V - \mathcal{P}_i = \mathrm{Pred}(\mathcal{P}_i)$ for $i \in \{2,4\}$, i.e. $\mathcal{P}_2$ and $\mathcal{P}_4$ are systems of pseudo-intents of $\mathcal{T}$. The non-redundant bases T_2 and T_4 of $\mathcal{T}$ given by $\mathcal{P}_2$ and $\mathcal{P}_4$ (see Theorem 1) are the following:

$$T_2 = \{\{{}^{0.5}/z\} \Rightarrow \{y, {}^{0.5}/z\}, \{y\} \Rightarrow \{y, {}^{0.5}/z\}\},$$
$$T_4 = \{\{{}^{0.5}/y, {}^{0.5}/z\} \Rightarrow \{y, {}^{0.5}/z\}, \{y\} \Rightarrow \{y, {}^{0.5}/z\}\}.$$

To sum up, for $\mathcal{T}$ there are two distinct systems of pseudo-intents.

In case of globalization on finite $\mathbf{L}$, for each data table with fuzzy attributes there is exactly one system of pseudo-intents, see [6], [9]. We now show that for globalization, $\mathcal{P}$ can be found using algorithm GBASIS (see Fig. 2) which uses the above-described graph procedure but takes advantage of a special nature of globalization: in order to find the basis, it suffices to traverse through the nodes of a graph in lexical order. The following lemma characterizes basic properties of GBASIS.

Lemma 3: Let * be globalization on a finite residuated lattice $\mathbf{L}$, $\mathcal{T} = \langle X, Y, I \rangle$ be a data table with fuzzy attributes, V be defined by (17), E be defined by (18), $\mathcal{S}$ be a sequence of lexically ordered elements of V. Whenever GBASIS reaches the beginning of the while-loop, the following conditions are true:

(i) if $V - \mathcal{Q} = \mathrm{Pred}(\mathcal{Q})$ then $\mathcal{P} \subseteq \mathcal{Q}$;
(ii) $\mathcal{S}$ consists of lexically ordered elements of $(V - \mathcal{P}) - \mathrm{Pred}(\mathcal{P})$.

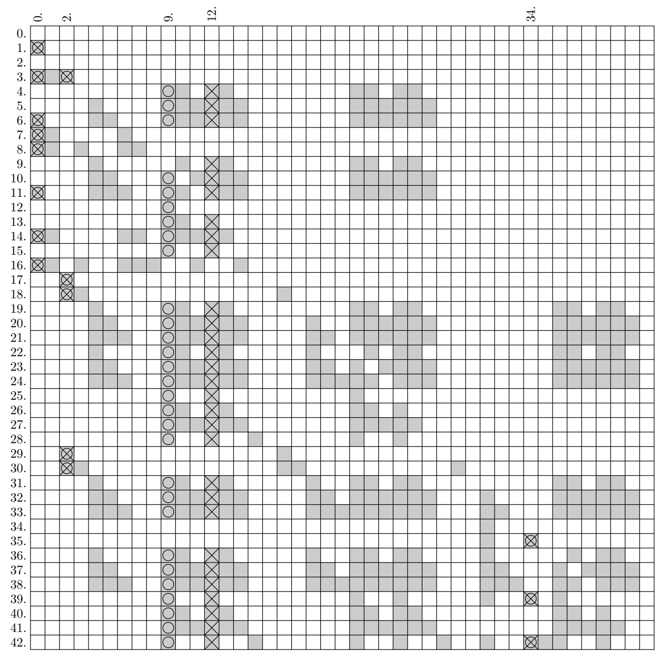
Proof: The proof is technically involved. Due to lack of space we postpone the proof to a forthcoming paper. ■

Now, we have the following characterization:

Theorem 3: If * is globalization on a finite residuated lattice then for each data table with fuzzy attributes there is exactly one system of pseudo-intents. Moreover, the system of pseudo-intents can be computed by GBASIS.

TABLE I
DATA TABLE WITH FUZZY ATTRIBUTES

	size		distance	
	small	large	far	near
Mercury	1	0	0	1
Venus	1	0	0	1
Earth	1	0	0	1
Mars	1	0	0.5	1
Jupiter	0	1	1	0.5
Saturn	0	1	1	0.5
Uranus	0.5	0.5	1	0
Neptune	0.5	0.5	1	0
Pluto	1	0	1	0



Proof: Let V be defined by (17), E be defined by (18), and $\mathcal{S}$ be a sequence of lexically ordered elements of V. Observe that for such $\mathcal{S}$, GBASIS terminates after finitely many steps with empty $\mathcal{S}$. We now prove existence and uniqueness of a system of pseudo-intents.

Existence: It suffices to show that $\mathcal{P}$ computed by GBASIS satisfies $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$. We check both inclusions. The computation terminates reaching the condition of while-loop for empty $\mathcal{S}$. Applying Lemma 3 (ii), we get $\emptyset = (V - \mathcal{P}) - \mathrm{Pred}(\mathcal{P})$. This gives, $V - \mathcal{P} \subseteq \mathrm{Pred}(\mathcal{P})$. The converse inclusion is also true: if $P \in \mathrm{Pred}(\mathcal{P})$ then there is $Q \in \mathcal{P}$ such that $\langle P, Q \rangle \in E$, i.e. $P \neq Q$ and $||Q \Rightarrow Q^{\downarrow\uparrow}||_P \neq 1$. That is, properties of globalization give $Q \subset P$, and $Q^{\downarrow\uparrow} \nsubseteq P$. Since $Q \in \mathcal{P}$, Q was added to $\mathcal{P}$ in a certain step of GBASIS, when Q was the first element of $\mathcal{S}'$. By Lemma 3 (ii), $\mathcal{S}'$ was lexically ordered. Since Q is lexically smaller than P (recall that $Q \subset P$, and lexical order is a linear order extending $\subset$), we have $P \notin \mathrm{UPDATE}(\mathcal{S}')$ which further gives $P \notin \mathcal{P}$, i.e. $P \in V - \mathcal{P}$. To sum up, we have shown $V - \mathcal{P} = \mathrm{Pred}(\mathcal{P})$, i.e. $\mathcal{P}$ is a system of pseudo-intents on account of Theorem 2.

Uniqueness: Let $\mathcal{Q}$ satisfy $V - \mathcal{Q} = \mathrm{Pred}(\mathcal{Q})$. By Lemma 3 (i), $\mathcal{P} \subseteq \mathcal{Q}$. Since both $\mathcal{P}$ and $\mathcal{Q}$ are maximal independent sets by Lemma 2, we have $\mathcal{P} = \mathcal{Q}$. ■

V. Examples and conclusions

Let $\mathbf{L}$ be a three-element Łukasiewicz chain with $L = \{0, 0.5, 1\}$. Consider a data table $\mathcal{T}$ given by Table I (top). The set X of object consists of objects "Mercury", "Venus", $\ldots$, Y contains four attributes: size of a planet (small / large), distance from the sun (far / near). For globalization on $\mathbf{L}$, there is a unique system of pseudo-intents [6] (cf. also Theorem 3) for $\mathcal{T}$ which induces the following basis T (see Theorem 1) of $\mathcal{T}$ (attributes are abbreviated by s—small, l—large, f—far, and n—near):

$$T = \{\{n\} \Rightarrow \{s,n\}, \{f, {}^{0.5}/n\} \Rightarrow \{l, f, {}^{0.5}/n\},$$
$$\{{}^{0.5}/l\} \Rightarrow \{{}^{0.5}/l, f\}, \{l, f\} \Rightarrow \{l, f, {}^{0.5}/n\},$$
$$\{{}^{0.5}/s, {}^{0.5}/n\} \Rightarrow \{s, n\}, \{s, {}^{0.5}/l, f\} \Rightarrow \{s, l, f, n\}\}.$$

If we replace globalization by identity, we obtain two distinct systems of pseudo-intents $\mathcal{P}_1$ and $\mathcal{P}_2$, both consist of four elements. Table I (bottom) contains the incidence matrix of relation $E \subseteq V \times V$ defined by (18). For brevity, the elements of V are denoted by numbers $0, \ldots, 42$. White box on a position P (row) and Q (column) indicates that $\langle P, Q\rangle \notin E$; gray box means $\langle P, Q\rangle \in E$, "$\bigcirc$" ("$\times$") indicates that $Q \in \mathcal{P}_1$ ($Q \in \mathcal{P}_2$) and $P \in \mathrm{Pred}(Q)$. Bases T_1 and T_2 given by $\mathcal{P}_1$ and $\mathcal{P}_2$ are the following:

$$T_1 = \{\{n\} \Rightarrow \{{}^{0.5}/s, n\}, \{f, {}^{0.5}/n\} \Rightarrow \{{}^{0.5}/l, f, {}^{0.5}/n\},$$
$$\{l\} \Rightarrow \{l, f, {}^{0.5}/n\},$$
$$\{s, {}^{0.5}/l, {}^{0.5}/f\} \Rightarrow \{s, {}^{0.5}/l, {}^{0.5}/f, {}^{0.5}/n\}\}.$$

$$T_2 = \{\{n\} \Rightarrow \{{}^{0.5}/s, n\}, \{f, {}^{0.5}/n\} \Rightarrow \{{}^{0.5}/l, f, {}^{0.5}/n\},$$
$$\{l, {}^{0.5}/f\} \Rightarrow \{l, f, {}^{0.5}/n\},$$
$$\{s, {}^{0.5}/l, {}^{0.5}/f\} \Rightarrow \{s, {}^{0.5}/l, {}^{0.5}/f, {}^{0.5}/n\}\}.$$

If $\mathbf{L}$ is a three-element Gödel chain with identity, we also have two distinct systems of pseudo-intents. The corresponding bases of $\mathcal{T}$ are the following:

$$T_1 = \{\{{}^{0.5}/l\} \Rightarrow \{{}^{0.5}/l, f\}, \{{}^{0.5}/l, f, {}^{0.5}/n\} \Rightarrow \{l, f, {}^{0.5}/n\},$$
$$\{{}^{0.5}/s, {}^{0.5}/n\} \Rightarrow \{s, n\}, \{{}^{0.5}/s, l, f\} \Rightarrow \{s, l, f\},$$
$$\{s, {}^{0.5}/l, f\} \Rightarrow \{s, l, f\}\}.$$

$$T_2 = \{\{{}^{0.5}/l, {}^{0.5}/f\} \Rightarrow \{{}^{0.5}/l, f\},$$
$$\{{}^{0.5}/l, f, {}^{0.5}/n\} \Rightarrow \{l, f, {}^{0.5}/n\}, \{l\} \Rightarrow \{l, f\},$$
$$\{{}^{0.5}/s, {}^{0.5}/n\} \Rightarrow \{s, n\}, \{{}^{0.5}/s, l, f\} \Rightarrow \{s, l, f\},$$
$$\{s, {}^{0.5}/l, f\} \Rightarrow \{s, l, f\}\}.$$

In this particular case, the systems of pseudo-intents have various sizes (5 and 6 elements). Thus, even though a basis defined by pseudo-intents is non-redundant (one cannot remove any implication without violating completeness), it is not minimal in terms of its size in general. Table II summarizes average counts and sizes of bases for randomly generated data tables with fuzzy attributes. The table on top contains results for sparse data tables (80% of table entries are zeros), the second table contains results for data tables with average density. Table rows indicate number of objects in randomly generated data tables; columns $\mathbf{L}_1$, $\mathbf{L}_2$, and $\mathbf{L}_3$ denote structures of truth degrees: three-element Łukasiewicz chain with globalization, three-element Łukasiewicz chain with identity, and three-element Gödel chain with identity, respectively. Column "count" contains average numbers of bases (for $\mathbf{L}_1$, "count" is omitted because there is always exactly one basis), column "size" contains average sizes of a bases. This experiment shows that sparse data tables usually lead to larger amount of bases and the bases themselves are greater than in case of tables with average density.

TABLE II

Average counts and sizes of bases

	$\mathbf{L}_1$	$\mathbf{L}_2$		$\mathbf{L}_3$	
rows	size	count	size	count	size
5	5.54	16.52	3.90	22.16	5.12
10	7.00	10.58	4.71	145.60	7.19
15	7.66	9.60	5.30	469.08	6.73
20	8.16	8.60	5.64	127.20	6.78
25	7.96	5.96	6.25	104.64	6.45
30	8.38	6.84	6.02	108.96	7.06
35	8.58	6.16	6.06	78.70	7.12
40	8.48	5.26	5.79	43.70	7.21

	$\mathbf{L}_1$	$\mathbf{L}_2$		$\mathbf{L}_3$	
rows	size	count	size	count	size
5	7.26	3.78	4.73	6.78	6.94
10	8.74	1.84	5.77	11.66	7.46
15	9.00	1.15	5.79	7.04	7.23
20	8.11	1.15	5.24	4.88	5.45
25	7.66	1.04	4.73	4.12	5.51
30	7.16	1.02	4.25	1.66	4.12
35	6.90	1.04	4.17	1.56	4.51
40	6.42	1.00	3.96	1.06	3.73

VI. Acknowledgment

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] Abiteboul S. *et al.*: The Lowell database research self-assessment. *Communications of ACM* **48**(5)(2005), 111–118.

[2] Bělohlávek R.: Similarity relations in concept lattices. *J. Logic Comput.* 10(6):823–845, 2000.

[3] Bělohlávek R.: Granulation and granularity via conceptual structures: A perspective from the point of view of fuzzy concept lattices. In: Lin T.Y., Yao Y.Y., Zadeh L.(eds.): *Data Mining, Rough Sets, and Granular Computing.* Physica-Verlag, Heidelberg, New York, 2002, pp. 265–289.

[4] Bělohlávek R.: *Fuzzy Relational Systems: Foundations and Principles.* Kluwer, Academic/Plenum Publishers, New York, 2002.

[5] Bělohlávek R.: Concept lattices and order in fuzzy logic. *Ann. Pure Appl. Logic* **128**(2004), 277–298.

[6] Bělohlávek R., Chlupová M., Vychodil V.: Implications from data with fuzzy attributes. AISTA 2004 in Cooperation with the IEEE Computer Society Proceedings, 2004, 5 pages, ISBN 2–9599776–8–8.

[7] Bělohlávek R., Vychodil V.: Implications from data with fuzzy attributes vs. scaled binary attributes. In: FUZZ-IEEE 2005, The IEEE International Conference on Fuzzy Systems, May 22–25, 2005, Reno (Nevada, USA), pp. 1050–1055 (proceedings on CD), abstract in printed proceedings, p. 53, ISBN 0–7803–9158–6.

[8] Bělohlávek R., Vychodil V.: Reducing the size of fuzzy concept lattices by hedges. In: FUZZ-IEEE 2005, The IEEE International Conference on Fuzzy Systems, May 22–25, 2005, Reno (Nevada, USA), pp. 663–668 (proceedings on CD), abstract in printed proceedings, p. 44, ISBN 0–7803–9158–6.

[9] Bělohlávek R., Vychodil V.: Proc. IFSA 2005 World Congress, July 28-31, 2005, Beijing, China, Vol. I, pp. 622–627, Springer, ISBN 7–302–11377–7.

[10] Bělohlávek R., Vychodil V.: Fuzzy attribute implications: computing non-redundant bases using maximal independent sets. Poster at AI 2005, Sydney, Australia.

[11] Ganter B., Wille R.: *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin, 1999.

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

[13] Guigues J.-L., Duquenne V.: Familles minimales d'implications informatives resultant d'un tableau de données binaires. *Math. Sci. Humaines* **95**(1986), 5–18.
[14] Hájek P.: *Metamathematics of Fuzzy Logic.* Kluwer, Dordrecht, 1998.
[15] Hájek P.: On very true. *Fuzzy Sets and Systems* **124**(2001), 329–333.
[16] Johnson D. S., Yannakakis M., Papadimitrou C. H.: On generating all maximal independent sets. *Inf. Processing Letters* **15**(1988), 129–133.
[17] Klir G. J., Yuan B.: *Fuzzy Sets and Fuzzy Logic. Theory and Applications.* Prentice Hall, 1995.
[18] Maier D.: *The Theory of Relational Databases.* Computer Science Press, Rockville, 1983.
[19] Takeuti G., Titani S.: Globalization of intuitionistic set theory. *Annals of Pure and Applied Logic* **33**(1987), 195–211.