

BASIC ALGORITHM FOR ATTRIBUTE IMPLICATIONS AND FUNCTIONAL DEPENDENCIES IN GRADED SETTING

RADIM BELOHLAVEK

*Dept. Systems Science & Industrial Eng., Binghamton University – SUNY
Binghamton, NY, 13902, U. S. A., e-mail: rbelohla@binghamton.edu*

and

*Department of Computer Science, Palacky University, Olomouc
Tomkova 40, CZ-779 00 Olomouc, Czech Republic*

and

VILEM VYCHODIL

*Dept. Systems Science & Industrial Eng., Binghamton University – SUNY
Binghamton, NY, 13902, U. S. A., e-mail: vychodil@binghamton.edu*

and

*Department of Computer Science, Palacky University, Olomouc
Tomkova 40, CZ-779 00 Olomouc, Czech Republic*

Received (received date)

Revised (revised date)

Communicated by Editor's name

ABSTRACT

We present GLINCLOSURE, a graded extension of the well-known LINCLOSURE algorithm. GLINCLOSURE can be used to compute degrees of semantic entailment from sets of fuzzy attribute implications. It can also be used together with graded extension of Ganter's NEXTCLOSURE algorithm to compute non-redundant bases of data tables with fuzzy attributes. We present foundations, the algorithm, analysis of its complexity, implementation details, and illustrative examples.

Keywords: Attribute implication; Functional dependence; Graded attributes.

1. Introduction

Fuzzy logic is a formal framework for dealing with a particular type of imprecision called vagueness. A key idea of fuzzy logic is a graded approach to truth in which we allow for truth degrees other than 0 (falsity) and 1 (full truth). This enables us to consider truth of propositions to various degrees, e.g., proposition “Peter is old” can be assigned a degree 0.8, indicating that “Peter is old” is almost (fully) true. One way of looking at the proposition “Peter is old” being true to degree 0.8 is that it expresses a *graded attribute* “being old to degree 0.8” of the *object* “Peter”. When dealing with multiple graded attributes, we often need to determine

their dependencies. In [2, 5] we have introduced fuzzy attribute implications as particular dependencies between graded attributes in data sets representing objects and their graded attributes (so-called data tables with fuzzy attributes). A fuzzy attribute implication can be seen as a rule of the form “ $A \Rightarrow B$ ”, saying “for each object from the data set: if the object has all graded attributes from A , then it has all graded attributes from B ”. We have proposed several ways to compute, given an input data set represented by a data tables with fuzzy attributes, a minimal set of dependencies describing all dependencies which are valid (true) in the table, see [7] for a survey.

In this paper we focus on computational aspects of one of the algorithms proposed so far. Namely, we show how to compute fixed points of certain fuzzy closure operators that appear in algorithms from [2, 11]. We introduce an extended version of the LINCLOSURE algorithm which is well known from database systems [23]. Compared to the original LINCLOSURE, our extended algorithm, called a GRADED LINCLOSURE (shortly, a GLINCLOSURE) is more versatile (this is discussed in Section 3) while having the same asymptotic time complexity as LINCLOSURE. This is an important feature since it is often the case that fuzzy logic extensions of classical algorithms proposed in the literature are of significantly higher time complexity than their classical counterparts. Since there is a close relationship between dependencies in data tables with fuzzy attributes and data tables over domains with similarity relations [8, 9], one can also use GLINCLOSURE for computing functional dependencies in data tables over domains with similarity relations. The latter naturally appear in an extension of Codd’s relational model which takes into account similarities on domains, see [8, 9].

2. Preliminaries and Motivation

In this section we present preliminaries of fuzzy logic, basic notions of fuzzy attribute implications, and several new concepts which will be used in further sections. More details can be found in [1, 17, 19, 22, 24] and [2, 5, 7]. In Section 2.2 we also present motivations for developing LINCLOSURE in fuzzy setting.

2.1. Fuzzy Logic and Fuzzy Set Theory

Since fuzzy logic and fuzzy sets are developed using general structures of truth degrees, we first introduce structures of truth degrees which are used in our approach. Our basic structures of truth degrees will be so-called complete residuated lattices with hedges, see [1, 17, 19, 20, 24]. A complete residuated lattice with hedge is an algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, *, 0, 1 \rangle$ such that

- (i) $\langle L, \wedge, \vee, 0, 1 \rangle$ is a complete lattice with 0 and 1 being the least and greatest element of L , respectively;
- (ii) $\langle L, \otimes, 1 \rangle$ is a commutative monoid (i.e., \otimes is commutative, associative, and for each $a \in L$ we have $a \otimes 1 = 1 \otimes a = a$);

(iii) \otimes and \rightarrow satisfy so-called adjointness property:

$$a \otimes b \leq c \quad \text{iff} \quad a \leq b \rightarrow c \quad (1)$$

for each $a, b, c \in L$;

(iv) hedge $*$ is a unary operation $*$: $L \rightarrow L$ satisfying, for each $a, b \in L$,

$$1^* = 1, \quad (2)$$

$$a^* \leq a, \quad (3)$$

$$(a \rightarrow b)^* \leq a^* \rightarrow b^*, \quad (4)$$

$$a^{**} = a^*. \quad (5)$$

Each $a \in L$ is called a truth degree. \otimes and \rightarrow are (truth functions of) “fuzzy conjunction” and “fuzzy implication”. Hedge $*$ is a (truth function of) logical connective “very true” and properties of hedges have natural interpretations [19, 20].

A common choice of \mathbf{L} is a structure with $L = [0, 1]$ (real 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 $[0, 1]$ are:

(i) Łukasiewicz: $a \otimes b = \max(a + b - 1, 0)$; $a \rightarrow b = \min(1 - a + b, 1)$;

(ii) Gödel: $a \otimes b = \min(a, b)$; $a \rightarrow b = 1$ if $a \leq b$, $a \rightarrow b = b$ else;

(iii) Goguen (product): $a \otimes b = a \cdot b$; $a \rightarrow b = 1$ if $a \leq b$, $a \rightarrow b = \frac{b}{a}$ else.

Complete residuated lattices include also finite structures of truth degrees. For instance, one can put $L = \{a_0 = 0, a_1, \dots, a_n = 1\} \subseteq [0, 1]$ ($a_0 < \dots < a_n$) 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 . 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). Two boundary cases of hedges are

(i) identity, i.e. $a^* = a$ ($a \in L$);

(ii) so-called globalization [26]:

$$a^* = \begin{cases} 1 & \text{if } a = 1, \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Moreover, for each \mathbf{L} we consider a derived truth function \ominus defined by

$$a \ominus b = a \otimes ((a \rightarrow b)^* \rightarrow 0). \quad (7)$$

For $*$ being globalization, we have the following assertion:

Lemma 1 *Let $*$ be globalization, i.e. let $*$ be defined by (6). Then*

$$a \ominus b = \begin{cases} 0 & \text{if } a \leq b, \\ a & \text{else.} \end{cases} \quad (8)$$

Proof. If $*$ is globalization, we have $(a \rightarrow b)^* \in \{0, 1\}$. Moreover, it follows that $(a \rightarrow b)^* = 1$ if $a \leq b$; and $(a \rightarrow b)^* = 0$ if $a \not\leq b$. \square

Remark 1 Note that the derived truth function \ominus can be seen as a particular subtraction of truth degrees because $a \otimes ((a \rightarrow b)^* \rightarrow 0)$ can be described as a degree to which “ a is true and it is not (very) true that b is greater than a ”. The meaning of \ominus as a type of subtraction is apparent especially in case of globalization, see (8). Due to (8), the result of $a \ominus b$ is 0 if b is greater than a . We will comment on the purpose of \ominus later on.

Until otherwise mentioned, we assume that \mathbf{L} denotes a complete residuated lattice (with hedge $*$) which serves as a structure of truth degrees. Using \mathbf{L} , we define the following notions. An \mathbf{L} -set (a fuzzy set) A in universe U is a mapping $A: U \rightarrow L$, $A(u)$ being interpreted as “the degree to which u belongs to A ”. If U is a finite universe $U = \{u_1, \dots, u_n\}$ then an \mathbf{L} -set A in U can be denoted by

$$A = \{a_1/u_1, \dots, a_n/u_n\}, \quad (9)$$

meaning that $A(u_i)$ equals a_i ($i = 1, \dots, n$). 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 . Denote by $|A|$ the cardinality of the support set of A , i.e. $|A| = |\{u \in U \mid A(u) > 0\}|$. The operations with \mathbf{L} -sets are defined componentwise. For instance, the union of \mathbf{L} -sets $A, B \in \mathbf{L}^U$ is an \mathbf{L} -set $A \cup B$ in U such that $(A \cup B)(u) = A(u) \vee B(u)$ ($u \in U$). Due to Lemma 1, for $*$ being globalization, we get

$$(A \ominus B)(u) = \begin{cases} 0 & \text{if } A(u) \leq B(u), \\ A(u) & \text{else.} \end{cases} \quad (10)$$

Remark 2 Fuzzy set $A \ominus B$ can be interpreted as follows. A degree $(A \ominus B)(u)$ to which $u \in U$ belongs to $A \ominus B$ is a truth degree to which “ u belongs to A and u does not belong to B at least to which it belongs to A ”. Think of A as of a fuzzy set assigning to each $u \in U$ a threshold degree $A(u)$. Then if $B(u)$ exceeds the threshold given by $A(u)$, element u will not be present in the resulting fuzzy set $A \ominus B$ (it means that $(A \ominus B)(u) = 0$), i.e. “ u will be removed”. If $B(u)$ does not exceed the threshold, we have $(A \ominus B)(u) = A(u)$, i.e. “ u will be preserved”. The situation is depicted in Fig. 1 which shows fuzzy set A (solid line), fuzzy set B (dotted line) and $A \ominus B$ (bold line).

For $a \in L$ and $A \in \mathbf{L}^U$, we define an \mathbf{L} -set $a \otimes A$ (a -multiple of A) by

$$(a \otimes A)(u) = a \otimes A(u), \quad (11)$$

for each $u \in U$. Binary \mathbf{L} -relations (binary fuzzy relations) between U and V can be thought of as \mathbf{L} -sets in $U \times V$. Given $A, B \in \mathbf{L}^U$, we define $S(A, B) \in L$ by

$$S(A, B) = \bigwedge_{u \in U} (A(u) \rightarrow B(u)). \quad (12)$$

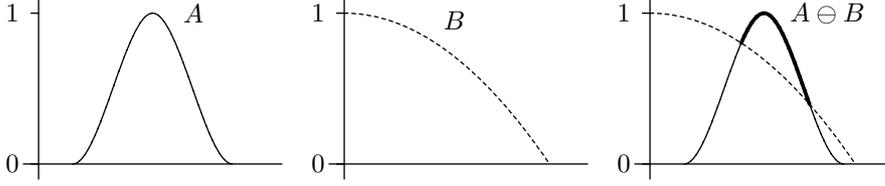


Figure 1: Subtraction of fuzzy sets

$S(A, B)$ is called a subsethood degree of A in B and it generalizes the classical subsethood relation \subseteq in a fuzzy setting. In particular, if \mathbf{L} (structure of truth degrees) is $\mathbf{2}$ (two-element Boolean algebra), then $\mathbf{2}$ -sets coincide in an obvious manner with (characteristic functions of) ordinary sets. Also, in case of $\mathbf{L} = \mathbf{2}$ we have that $S(A, B) = 1$ iff $A \subseteq B$. For general \mathbf{L} , we write $A \subseteq B$ iff $S(A, B) = 1$; and $A \subset B$ iff $S(A, B) = 1$ and $A \neq B$. As a consequence of properties of \rightarrow and \wedge , we get that $A \subseteq B$ iff $A(u) \leq B(u)$ for each $u \in U$, see [1, 17, 22].

A fuzzy closure operator with hedge $*$ (shortly, a fuzzy closure operator) [3] on a set U is a mapping $C: \mathbf{L}^U \rightarrow \mathbf{L}^U$ satisfying, for each $A, B_1, B_2 \in \mathbf{L}^U$,

$$A \subseteq C(A), \quad (13)$$

$$S(B_1, B_2)^* \leq S(C(B_1), C(B_2)), \quad (14)$$

$$C(A) = C(C(A)). \quad (15)$$

2.2. Fuzzy Attribute Implications

Let Y denote a *finite set of attributes*. Each \mathbf{L} -set $M \in \mathbf{L}^Y$ of attributes can be seen as a set of graded attributes because M prescribes, for each attribute $y \in Y$, a degree $M(y) \in L$. A *fuzzy attribute implication (over attributes Y)* is an expression $A \Rightarrow B$, where $A, B \in \mathbf{L}^Y$ are fuzzy sets of attributes. Fuzzy attribute implications (FAIs) represent particular data dependencies. The intuitive meaning we wish to give to $A \Rightarrow B$ is: “if it is (very) true that an object has all (graded) attributes from A , then it has also all (graded) attributes from B ”. Formally, for an \mathbf{L} -set $M \in \mathbf{L}^Y$ of attributes, we define a *truth degree* $\|A \Rightarrow B\|_M \in L$ to which $A \Rightarrow B$ is true in M by

$$\|A \Rightarrow B\|_M = S(A, M)^* \rightarrow S(B, M), \quad (16)$$

with $S(\cdot \cdot \cdot)$ defined by (12). 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 $A \Rightarrow B$. Note also that the hedge $*$ present in (16) serves as a modifier of interpretation of $A \Rightarrow B$ and plays an important technical role. If $*$ is globalization, i.e. if $*$ is

defined by (6), then $\|A \Rightarrow B\|_M = 1$ (i.e., $A \Rightarrow B$ is fully true in M) iff we have:

$$\text{if } A \subseteq M, \text{ then } B \subseteq M. \quad (17)$$

More information about the role of hedges can be found in [2, 5, 7]. See also [25] for a related approach to attribute implications in fuzzy setting.

Let T be a set of fuzzy attribute implications. An \mathbf{L} -set $M \in \mathbf{L}^Y$ is called a *model of T* if, for each $A \Rightarrow B \in T$, $\|A \Rightarrow B\|_M = 1$. The set of all models of T will be denoted by $\text{Mod}(T)$, i.e.

$$\text{Mod}(T) = \{M \in \mathbf{L}^Y \mid \text{for each } A \Rightarrow B \in T: \|A \Rightarrow B\|_M = 1\}. \quad (18)$$

A *degree* $\|A \Rightarrow B\|_T$ to which $A \Rightarrow B$ semantically follows from T is defined by

$$\|A \Rightarrow B\|_T = \bigwedge_{M \in \text{Mod}(T)} \|A \Rightarrow B\|_M. \quad (19)$$

Described verbally, $\|A \Rightarrow B\|_T$ is defined to be the degree to which “ $A \Rightarrow B$ is true in each model of T ”. Hence, degrees $\|\cdot\|_T$ defined by (19) represent degrees of semantic entailment from T . Let us note that degrees $\|\cdot\|_T$ can also be fully described via the (syntactic) concept of a *provability degree*, see [7, 12].

The set $\text{Mod}(T)$ of all models of T form a particular fuzzy closure system in Y , see [11] for details. Thus, for each \mathbf{L} -set $M \in \mathbf{L}^Y$ we can consider its closure in $\text{Mod}(T)$ which is then the least model of T containing M . The closure operator associated with $\text{Mod}(T)$ can be described as follows. First, for any set T of FAIs and any \mathbf{L} -set $M \in \mathbf{L}^Y$ of attributes define an \mathbf{L} -set $M^T \in \mathbf{L}^Y$ of attributes by

$$M^T = M \cup \bigcup \{S(A, M)^* \otimes B \mid A \Rightarrow B \in T\}. \quad (20)$$

Note that if $*$ is globalization, (20) simplifies as follows:

$$M^T = M \cup \bigcup \{B \mid A \Rightarrow B \in T \text{ and } A \subseteq M\}. \quad (21)$$

Using (20), for each $n \in \mathbb{N}_0$ we define a fuzzy set $M^{T_n} \in \mathbf{L}^Y$ of attributes by

$$M^{T_n} = \begin{cases} M & \text{for } n = 0 \\ (M^{T_{n-1}})^T & \text{for } n \geq 1. \end{cases} \quad (22)$$

Finally, we define an operator $cl_T: \mathbf{L}^Y \rightarrow \mathbf{L}^Y$ by

$$cl_T(M) = \bigcup_{n=0}^{\infty} M^{T_n}. \quad (23)$$

The following assertion shows the importance of cl_T .

Theorem 1 (see [11]) *Let \mathbf{L} be a finite residuated lattice with hedge, T be a set of fuzzy attribute implications. Then*

- (i) cl_T defined by (23) is a fuzzy closure operator;
- (ii) $cl_T(M)$ is the least model of T containing M , i.e. $cl_T(M) \in \text{Mod}(T)$ and, for each $N \in \text{Mod}(T)$, if $M \subseteq N$ then $cl_T(M) \subseteq N$;

(iii) $\|A \Rightarrow B\|_T = S(B, cl_T(A))$. \square

Remark 3 Note that Theorem 1 (iii) says that degrees of semantic entailment from sets of fuzzy attribute implications can be expressed as subethood degrees of consequents of FAIs into least models generated by antecedents of FAIs. Hence, a single model of T suffices to express the degree $\|A \Rightarrow B\|_T$, cf. definition (19). In other words, an efficient procedure for computing of closures $cl_T(\dots)$ would give us an efficient procedure to compute degrees of semantic entailment.

Another area in which a closure operator similar to (23) appears is the computation of non-redundant bases of data tables with fuzzy attributes. A *data table with fuzzy attributes* is a triplet $\langle X, Y, I \rangle$ where X is a set of objects, Y is a finite set of attributes (the same as above), 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 “object x has attribute y ”. $\langle X, Y, I \rangle$ can be thought of 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 $\langle X, Y, I \rangle$ corresponding to an object $x \in X$ can be seen as a set I_x of graded attributes (a fuzzy set of attributes) to which an attribute $y \in Y$ belongs to a degree $I_x(y) = I(x, y)$. Furthermore, a *degree* $\|A \Rightarrow B\|_{\langle X, Y, I \rangle}$ to which $A \Rightarrow B$ is true in data table $\langle X, Y, I \rangle$ is defined by

$$\|A \Rightarrow B\|_{\langle X, Y, I \rangle} = \bigwedge_{x \in X} \|A \Rightarrow B\|_{I_x}. \quad (24)$$

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

In order to describe particular non-redundant bases of data tables with fuzzy attributes we need to recall basic notions of formal concept analysis of data tables with fuzzy attributes [4, 7]. Given a data table $\langle X, Y, I \rangle$, for $A \in \mathbf{L}^X$, $B \in \mathbf{L}^Y$ we define $A^\uparrow \in \mathbf{L}^Y$ and $B^\downarrow \in \mathbf{L}^X$ by

$$A^\uparrow(y) = \bigwedge_{x \in X} (A(x)^* \rightarrow I(x, y)), \quad (25)$$

$$B^\downarrow(x) = \bigwedge_{y \in Y} (B(y) \rightarrow I(x, y)). \quad (26)$$

Operators $^\downarrow, ^\uparrow$ form so-called Galois connection with hedge, see [4]. The set of all fixed points of $^\downarrow, ^\uparrow$ (so-called fuzzy concepts) hierarchically ordered by a subconcept-superconcept relation is called a *fuzzy concept lattice with hedge*, see [4, 7]. A crucial role in determining a non-redundant basis of a given $\langle X, Y, I \rangle$ is played by an operator which is a modification of cl_T , see (23). The modified operator can be

described as follows. For $M \in \mathbf{L}^Y$ put

$$M^{T^*} = M \cup \bigcup \{S(A, M)^* \otimes B \mid A \Rightarrow B \in T \text{ and } A \neq M\}. \quad (27)$$

If $*$ is globalization, (27) is equivalent to

$$M^{T^*} = M \cup \bigcup \{B \mid A \Rightarrow B \in T \text{ and } A \subset M\}. \quad (28)$$

We can now define an operator cl_{T^*} in much the same way as cl_T :

$$M^{T_n^*} = \begin{cases} M & \text{for } n = 0 \\ (M^{T_{n-1}^*})^{T^*} & \text{for } n \geq 1, \end{cases} \quad (29)$$

$$cl_{T^*}(M) = \bigcup_{n=0}^{\infty} M^{T_n^*}. \quad (30)$$

For cl_{T^*} defined by (29), we have the following

Theorem 2 (see [2, 5, 7]) *Let \mathbf{L} be a finite residuated lattice with globalization, $\langle X, Y, I \rangle$ be a data table with fuzzy attributes. Then there is T such that*

- (i) cl_{T^*} is a fuzzy closure operator;
- (ii) a set of FAIs defined by $\{P \Rightarrow P^{\downarrow\uparrow} \mid P = cl_{T^*}(P) \text{ and } P \neq P^{\downarrow\uparrow}\}$ is a non-redundant basis of $\langle X, Y, I \rangle$. □

Remark 4 From Theorem 2 it follows that for $*$ being globalization a non-redundant basis of $\langle X, Y, I \rangle$ is determined by particular fixed points of cl_{T^*} , namely, by fuzzy sets $P \in \mathbf{L}^Y$ of attributes such that $P = cl_{T^*}(P)$ and $P \neq P^{\downarrow\uparrow}$. The basis given by Theorem 2 is also a minimal one, i.e. each set of fuzzy attribute implications which is complete in $\langle X, Y, I \rangle$ has at least the same number of FAIs as the basis given by Theorem 2, see [5]. Notice that we have not specified the set T of fuzzy attribute implications which is used by cl_{T^*} . A detailed description of that set is outside the scope of our paper, see [2, 5, 7]. An approach for general hedges has been presented in [6, 10]. Let us just mention that T is computationally tractable. Fuzzy sets of attributes satisfying $P = cl_{T^*}(P)$ and $P \neq P^{\downarrow\uparrow}$ will occasionally be referred to as *pseudo-intents* or *pseudo-closed fuzzy set of attributes*. This is for the sake of consistency with [2, 5, 7], cf. also [14, 15, 16, 18].

3. Graded LinClosure

Throughout this section, we assume that \mathbf{L} is a finite linearly ordered residuated lattice with globalization, see (6). Structure \mathbf{L} represents a finite linear scale of truth degrees.

Problem Setting Given a fuzzy set $M \in \mathbf{L}^Y$ of attributes we wish to compute its closures $cl_T(M)$ and $cl_{T^*}(M)$ defined by (23) and (30), respectively. First, note that cl_T and cl_{T^*} differ only in non-strict/strict fuzzy set inclusions “ \subseteq ” and “ \subset ” used in (21) and (28). A direct method (naive algorithm) to compute $cl_T(M)$ and $cl_{T^*}(M)$, which is given by the definitions of cl_T and cl_{T^*} , leads to an algorithm similar to CLOSURE which is known from database systems [23]. In more detail:

for a given M , we iterate through all FAIs in T and for each $A \Rightarrow B \in T$ we test if $A \subseteq M$ ($A \subset M$); if so, we add B to M (i.e., we set $M := M \cup B$) and repeat the process until M cannot be enlarged; the resulting M is the closure under cl_T (cl_{T^*}) of the original fuzzy set M . Clearly, this procedure is sound. Let n be the number of attributes in Y and p be the number of FAIs in T . In the worst case, we have to make p^2 iterations in order to compute the closure because in each loop through all FAIs in T there can be only one $A \Rightarrow B$ such that $A \subseteq M$ ($A \subset M$) and $B \not\subseteq M$ (i.e., only one FAI from T causes M to be enlarged). Moreover, for each $A \Rightarrow B$ we need n steps to check the non-strict/strict subsethood $A \subseteq M$ ($A \subset M$). To sum up, the complexity of this algorithm is $O(np^2)$, where n is the number of attributes and p is the number of FAIs from T (cf. [23]).

In this section we present an improved version of the algorithm, so-called GLINCLOSURE (GRADED LINCLOSURE), which computes $cl_T(M)$ and $cl_{T^*}(M)$ with complexity $O(n)$, where n is the size of the input. GLINCLOSURE uses each FAI from T only once and allows us to check the non-strict/strict inclusions $A \subseteq M$ ($A \subset M$) which appear in (21) and (28) in a constant time. Our algorithm results by extending LINCLOSURE [23] so that

- (i) we can use *fuzzy sets of attributes* instead of classical sets (this brings new technical problems with efficient comparing of truth degrees and checking of strict inclusion, see below);
- (ii) we can use the algorithm also to compute *systems of pseudo-intents* (i.e., fixed points of cl_{T^*}), and thus non-redundant bases (the original LINCLOSURE [23] cannot be used to compute pseudo-intents [16], it can only compute fixed points of the classical counterpart of cl_T), this also brings technical complications since we have to maintain a “waitlist of attributes” which can possibly be updated (or not) in future iterations.

In what follows we present a detailed description of the algorithm and analysis of its complexity.

Input and Output of the Algorithm The input for GLINCLOSURE consists of a set T of fuzzy attribute implications over Y , a fuzzy set $M \in \mathbf{L}^Y$ of attributes, and a flag $PCLOSED \in \{false, true\}$. The meaning of $PCLOSED$ is the following. If $PCLOSED$ is set to *true*, the output of GLINCLOSURE is $cl_{T^*}(M)$ (the least fixed point of cl_{T^*} which contains M); if $PCLOSED$ is set to *false*, the output of GLINCLOSURE is $cl_T(M)$ (the least fixed point of cl_T , i.e. the least model of T , which contains M).

Representation of Graded Attributes During the computation, we represent fuzzy sets (**L**-sets) of attributes in Y by ordinary sets of tuples $\langle y, a \rangle$, where $y \in Y$ and $a \in L$. Namely, a fuzzy set $\{a_1/y_1, a_2/y_2, \dots, a_n/y_n\}$ ($a_1 \neq 0, \dots, a_n \neq 0$) will be represented by an ordinary set $\{\langle y_1, a_1 \rangle, \langle y_2, a_2 \rangle, \dots, \langle y_n, a_n \rangle\}$. We will use both notations $A(y) = a$ and $\langle y, a \rangle \in A$. Whenever we consider $\langle y, a \rangle$, we assume $a \neq 0$. From the implementational point of view, we may represent fuzzy sets of attributes

by lists of tuples $\langle y, a \rangle$ instead of sets of such tuples. In such a case, we write $(\langle y_1, a_1 \rangle, \langle y_2, a_2 \rangle, \dots, \langle y_n, a_n \rangle)$ instead of $\{\langle y_1, a_1 \rangle, \langle y_2, a_2 \rangle, \dots, \langle y_n, a_n \rangle\}$.

Quick Test of Subsethood We avoid repeated testing of inclusions in (21) and (28) analogously as in the original LINCLOSURE. For each fuzzy attribute implication $A \Rightarrow B$ from T we keep a record of the number of attributes due to which A is not contained in the constructed closure. If this number reaches zero, we get that $A \subseteq M$ and we can process $A \Rightarrow B$, see (21). The counter suffices to check non-strict subsethood which is needed to compute fixed points of cl_T . In order to check strict subsethood which is needed to compute $cl_{T^*}(M)$, we need to have a quick test to decide if $A \subset M$ provided we already know that $A \subseteq M$. The test can be done with the following notion of *cardinality* of fuzzy sets. Take a fixed monotone injective mapping $f_{\mathbf{L}} : L \rightarrow [0, 1]$. That is, $f_{\mathbf{L}}$ is injective, and for each $a, b \in L$, if $a \leq b$ then $f_{\mathbf{L}}(a) \leq f_{\mathbf{L}}(b)$. For each fuzzy set $M \in \mathbf{L}^Y$ of attributes we define a number $\text{card}(M) \in [0, \infty)$ by

$$\text{card}(M) = \sum_{\langle y, a \rangle \in M} f_{\mathbf{L}}(a). \quad (31)$$

For instance, if L is a subset of $[0, 1]$, we can put $f_{\mathbf{L}}(a) = a$ ($a \in L$), and thus $\text{card}(M) = \sum_{\langle y, a \rangle \in M} a$.

Lemma 2 *Let $A, B \in \mathbf{L}^Y$ such that $A \subseteq B$. Then $A \subset B$ iff $\text{card}(A) < \text{card}(B)$.*

Proof. Since $A \subseteq B$, we have that $A(y) \leq B(y)$ is true for each $y \in Y$. Since $f_{\mathbf{L}}$ is monotone, we get that $f_{\mathbf{L}}(A(y)) \leq f_{\mathbf{L}}(B(y))$ is true for each $y \in Y$. From it follows that $\text{card}(A) \leq \text{card}(B)$. If in addition $A \subset B$ then there exists $y \in Y$ such that $A(y) < B(y)$. Due to monotony and injectivity of $f_{\mathbf{L}}$, for such a $y \in Y$ we have $f_{\mathbf{L}}(A(y)) < f_{\mathbf{L}}(B(y))$. This immediately gives $\text{card}(A) < \text{card}(B)$. Conversely, suppose we have $\text{card}(A) < \text{card}(B)$. Since $A \subseteq B$, there must be $y \in Y$ with $A(y) < B(y)$ otherwise we would have $\text{card}(A) = \text{card}(B)$. This means $A \subset B$. \square

Remark 5 Note that checking strict inclusion in fuzzy setting is more difficult than in the ordinary case where one can decide it simply by comparing numbers of elements in both sets. In fuzzy setting, we can have fuzzy sets which have the same number of elements belonging to the sets (to a non-zero degree) but the sets may not be equal. For instance, if $L = [0, 1]$, then $A = \{0.7/y, 0.5/z\}$ and $B = \{0.9/y, 0.5/z\}$ both contain two elements (to a non-zero degree), i.e. $|A| = |B| = 2$ (see preliminaries). Hence, the values of $|\cdot \cdot \cdot|$ alone are not sufficient to decide $A \subset B$ provided that $A \subseteq B$. This is why we have introduced “cardinalities” by (31).

Data Structures Used During the Computation:

NEWDEP is a fuzzy set of attributes which is the closure being constructed;

CARDND is the cardinality of *NEWDEP* given by (31);

COUNT[$A \Rightarrow B$] is a nonnegative integer indicating the number of attributes from A such that $A(y) > \text{NEWDEP}(y)$, *COUNT*[$A \Rightarrow B$] = 0 means that A is a subset of *NEWDEP*;

$CARD[A \Rightarrow B]$ is a number indicating cardinality of A , it is used to decide if A is strictly contained in $NEWDEP$ when $COUNT[A \Rightarrow B]$ reaches zero;

$UPDATE$ is a fuzzy set of attributes which are waiting for update;

$WAITLIST$ is a list of (pointers to) fuzzy sets of attributes which can be added to $NEWDEP$ as soon as $NEWDEP$ will increase its cardinality; this is necessary if $PCLOSED = true$, $WAITLIST$ is not used if $PCLOSED = false$;

$LIST[y]$ is an attribute-indexed collection of (pointers to) FAIs from T such that $A \Rightarrow B$ is referenced in $LIST[y]$ iff $A(y) > 0$;

$SKIP[y][A \Rightarrow B] \in \{false, true\}$ indicates whether attribute y has already been updated in $A \Rightarrow B$ ($SKIP[y][A \Rightarrow B] = true$) or not ($SKIP[y][A \Rightarrow B] = false$); $SKIP$ is necessary in graded setting to avoid updating an attribute twice which may result in incorrect values of $COUNT[A \Rightarrow B]$;

$DEGREE[y][A \Rightarrow B]$ represents a degree in which attribute y is contained in A ;

$()$ denotes the empty list.

Remark 6 Note that $NEWDEP$, $UPDATE$, $COUNT$, and $LIST$ play a similar role as in $LINCLOSURE$, cf. [23].

Description of the Algorithm and Complexity Analysis The algorithm is depicted in Fig. 2. Let us mention that the algorithm has two basic parts:

- (i) *initialization of data structures* (lines 1–16), and
- (ii) *main computation* (lines 17–38).

Denote by k the number of truth degrees, i.e. $k = |L|$. Denote by n the length of the input which is the sum of attributes which belong to left-hand sides of FAIs to non-zero degrees, i.e.:

$$n = \sum_{A \Rightarrow B \in T} |A|.$$

The initialization can be done with time complexity $O(n)$. Of course, the initialization depends on data structures we choose for representing $LIST$, $SKIP$, $DEGREE$, $COUNT$, and $CARD$. Clearly, the values of $COUNT[A \Rightarrow B]$ and $CARD[A \Rightarrow B]$ can be computed in $|A|$ steps. The addition of a FAI to $LIST[y]$ can be done in a constant time. Structures $DEGREE$ and $SKIP$ can be represented by two-dimensional tables or by a vector of lists indexed by truth degrees which appear in ascendant manner. The latter representation will allow us to maintain $O(n)$ during the computation. In next section we propose an efficient structure encompassing the information from $LIST$, $SKIP$, $DEGREE$, $COUNT$, and $CARD$ whose initialization takes $O(kn)$ steps. Since k (number of truth degrees) is a multiplicative constant (size of \mathbf{L} is fixed and does not depend on the length of the input), the initialization is linearly dependent on the length of the input, i.e. it is indeed in $O(n)$. Also note that the number of pairwise distinct truth degrees that appear in left-hand sides of FAIs is usually much smaller than the size of the input data.

In the second part (computation), each graded attribute $\langle y, a \rangle$ is considered at most once for update. This is ensured in a similar way as in the ordinary case,

Input: a set T of FAIs over Y , a fuzzy set $M \in \mathbf{L}^Y$ of attributes,
and a flag $PCLOSED \in \{false, true\}$
Output: $cl_{T^*}(M)$ if $PCLOSED = true$, or $cl_T(M)$ if $PCLOSED = false$

Initialization:

```

1  if  $M = \emptyset$  and  $PCLOSED = true$ :
2    return  $\emptyset$ 
3   $NEWDEP := M$ 
4  for each  $A \Rightarrow B \in T$ :
5    if  $A = \emptyset$ :
6       $NEWDEP := NEWDEP \cup B$ 
7    else:
8       $COUNT[A \Rightarrow B] := |A|$ 
9       $CARD[A \Rightarrow B] := \text{card}(A)$ 
10     for each  $\langle y, a \rangle \in A$ :
11       add  $A \Rightarrow B$  to  $LIST[y]$ 
12        $DEGREE[y][A \Rightarrow B] := a$ 
13        $SKIP[y][A \Rightarrow B] := false$ 
14   $UPDATE := NEWDEP$ 
15   $CARDND := \text{card}(NEWDEP)$ 
16   $WAITLIST := ()$ 

```

Computation:

```

17 while  $UPDATE \neq \emptyset$ :
18   choose  $\langle y, a \rangle \in UPDATE$ 
19    $UPDATE := UPDATE - \{\langle y, a \rangle\}$ 
20   for each  $A \Rightarrow B \in LIST[y]$  such that
21      $SKIP[y][A \Rightarrow B] = false$  and  $DEGREE[y][A \Rightarrow B] \leq a$ :
22      $SKIP[y][A \Rightarrow B] = true$ 
23      $COUNT[A \Rightarrow B] := COUNT[A \Rightarrow B] - 1$ 
24     if  $COUNT[A \Rightarrow B] = 0$  and
25       ( $PCLOSED = false$  or  $CARD[A \Rightarrow B] < CARDND$ ):
26        $ADD := B \ominus NEWDEP$ 
27        $CARDND := CARDND + \sum_{\langle y, a \rangle \in ADD} (f_{\mathbf{L}}(a) - f_{\mathbf{L}}(NEWDEP(y)))$ 
28        $NEWDEP := NEWDEP \cup ADD$ 
29        $UPDATE := UPDATE \cup ADD$ 
30       if  $PCLOSED = true$  and  $ADD \neq \emptyset$ :
31         while  $WAITLIST \neq ()$ :
32           choose  $B \in WAITLIST$ 
33           remove  $B$  from  $WAITLIST$ 
34            $ADD := B \ominus NEWDEP$ 
35            $CARDND := CARDND + \sum_{\langle y, a \rangle \in ADD} (f_{\mathbf{L}}(a) - f_{\mathbf{L}}(NEWDEP(y)))$ 
36            $NEWDEP := NEWDEP \cup ADD$ 
37            $UPDATE := UPDATE \cup ADD$ 
38       if  $COUNT[A \Rightarrow B] = 0$  and  $PCLOSED = true$  and
39          $CARD[A \Rightarrow B] = CARDND$ :
40         add  $B$  to  $WAITLIST$ 
41 return  $NEWDEP$ 

```

Figure 2: Graded LinClosure

see [23]. For each fuzzy attribute implication, the value of *COUNT* reaches zero at most once. Then, the computation which follows (lines 24–27 or lines 32–35) is linearly dependent on the size of the left-hand side of the processed fuzzy attribute implication. This again depends on the representation of fuzzy sets and operations with fuzzy sets. If we represent fuzzy sets by a list of pairs of the form $\langle y, a \rangle$ where $y \in Y$ and $a \in L$ which is moreover sorted by the attributes (in some fixed order), we can perform all necessary operations in linear time proportional to the size of the fuzzy set. Thus, using analogous arguments as in case of the original LINCLOSURE [23], we get that GLINCLOSURE works with asymptotic time complexity $O(n)$. Note that lines 28–37 are not present in the ordinary LINCLOSURE. This is because if we intend to compute fixed points of cl_{T^*} , a graded attribute can be scheduled for updating (added to *UPDATE*) only if we know that the left-hand side of FAI is strictly contained in *NEWDEP*. This is checked at line 36. The loop starting at line 29 checks whether there are some graded attributes waiting for update. The position of the loop in the program ensures that it is processed only after *NEWDEP* has been enlarged (see lines 24–27). Thus, setting *PCLOSED* to *true* really leads to computing fixed points of cl_{T^*} . The condition at line 20 can be checked efficiently if we have an efficient representation of *SKIP* and *DEGREE*. That is, given an attribute $y \in Y$ and a truth degree $a \in L$ we need to obtain all FAIs such that y belongs to their left-hand sides at most to degree a . This can be easily done if for each $y \in Y$ we have a list of FAIs ordered by truth degrees in an ascendant manner (see the discussion above related to initialization). A data representation of *SKIP* and *DEGREE* will be shown in next section.

Remark 7 If \mathbf{L} (our structure of truth degrees) is a two-element Boolean algebra, i.e. if $L = \{0, 1\}$, GLINCLOSURE with *PCLOSED* set to *false* produces the same results as LINCLOSURE [23] (the only difference is that our algorithm allows also for FAIs of the form $\{ \} \Rightarrow B$ whereas the original LINCLOSURE does not). From this point of view, GLINCLOSURE is a generalization of LINCLOSURE. GLINCLOSURE is more versatile (even in the ordinary case): GLINCLOSURE can be used to compute pseudo-intents (and thus a non-redundant basis of data tables with fuzzy attributes) which cannot be done with the original LINCLOSURE (without additional modifications).

4. Implementation Details, Examples, and Remarks

As mentioned before, the efficiency of an implementation of GLINCLOSURE is closely connected with data structures. The information contained in *LIST*, *SKIP*, *DEGREE*, *COUNT*, and *CARD* can be stored in a single efficient data structure. This structure, called a *T-structure*, is a particular attribute-indexed vector of lists of pointers to structures carrying values from *COUNT* and *CARD*. We illustrate the construction of a *T-structure* by an example. Consider a set T of FAIs which

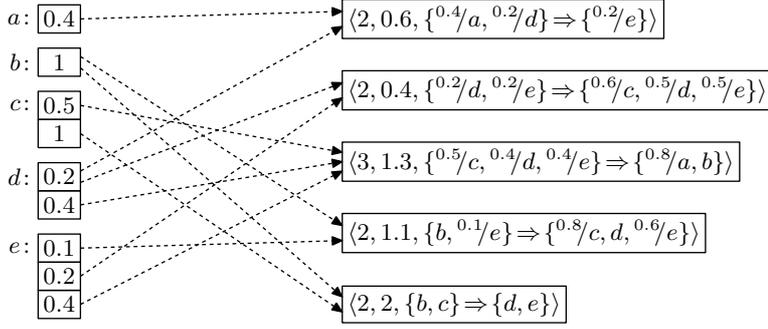


Figure 3: T -structure encompassing $LIST$, $SKIP$, $DEGREE$, $COUNT$, and $CARD$

consists of the following fuzzy attribute implications:

$$\begin{aligned}
 \varphi_1: \{ \} &\Rightarrow \{0.4/a, 0.1/d\}, & \varphi_4: \{0.5/c, 0.4/d, 0.4/e\} &\Rightarrow \{0.8/a, b\}, \\
 \varphi_2: \{0.4/a, 0.2/d\} &\Rightarrow \{0.2/e\}, & \varphi_5: \{b, 0.1/e\} &\Rightarrow \{0.8/c, d, 0.6/e\}, \\
 \varphi_3: \{0.2/d, 0.2/e\} &\Rightarrow \{0.6/c, 0.5/d, 0.5/e\}, & \varphi_6: \{b, c\} &\Rightarrow \{d, e\}.
 \end{aligned}$$

Since φ_1 is of the form $\{ \} \Rightarrow B$, its right-hand side is added to $NEWDEP$ and the implication itself is not contained in $LIST$ and other structures. The other formulas, i.e. $\varphi_2, \dots, \varphi_6$, are used to build a new T -structure which is depicted in Fig. 3. The T -structure can be seen as consisting of two main parts. First, a set of records encompassing information about the FAIs, $COUNT$, and $CARD$. For each FAI φ_i , we have a single record, called a T -record, of the form $\langle COUNT[\varphi_i], CARD[\varphi_i], \varphi_i \rangle$, see Fig. 3 (right). Second, an attribute-indexed vector of lists containing truth degrees and pointers to T -records, see Fig. 3 (left). A list which is indexed by attribute $y \in Y$ will be called a y -list. The aim of this part of the structure is to keep information about the occurrence of graded attributes that appear in left-hand sides of FAIs from T . In more detail, a y -list contains truth degree $a \in L$ iff there is at least one $A \Rightarrow B \in T$ such that $0 \neq A(y) = a$. Moreover, if a y -list contains a as its element, then it is connected via pointer to all T -records $\langle m, n, C \Rightarrow D \rangle$ such that $C(y) = a$. Because of the computational efficiency, each y -list is sorted by truth degrees in the ascendant manner. Note that pointers between elements of lists Fig. 3 (left) and T -records Fig. 3 (right) represent information in $SKIP$ ($SKIP[y][A \Rightarrow B] = false$ means that pointer from element $A(y)$ of y -list to T -record of $A \Rightarrow B$ is present). As one can see, a T -structure can be constructed by a sequential updating of the structure with time complexity $O(kn)$, where n is the size of the input (each graded attribute is considered once) and k is the number of truth degrees (this is an overhead needed to keep y -lists sorted). In the following examples, we will use a convenient notation for writing T -structures which correspond in an obvious way with graphs of the form of Fig. 3. For example, instead of Fig. 3, we write:

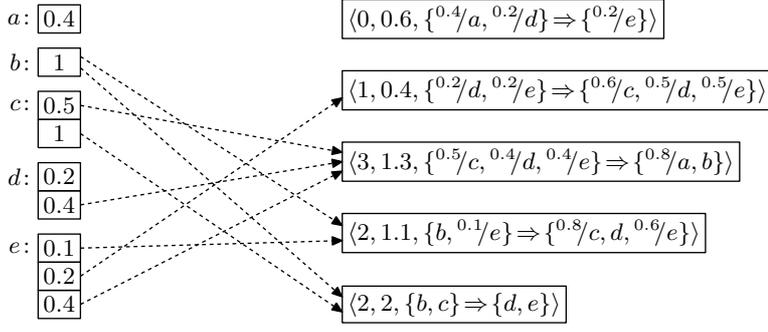


Figure 4: T -structure before processing the first FAI

a : $[(0.4, \langle 2, 0.6, \varphi_2 \rangle)]$
 b : $[(1, \langle 2, 2, \varphi_6 \rangle), \langle 2, 1.1, \varphi_5 \rangle]$
 c : $[(0.5, \langle 3, 1.3, \varphi_4 \rangle), (1, \langle 2, 2, \varphi_6 \rangle)]$
 d : $[(0.2, \langle 2, 0.4, \varphi_3 \rangle), \langle 2, 0.6, \varphi_2 \rangle], (0.4, \langle 3, 1.3, \varphi_4 \rangle)]$
 e : $[(0.1, \langle 2, 1.1, \varphi_5 \rangle), (0.2, \langle 2, 0.4, \varphi_3 \rangle), (0.4, \langle 3, 1.3, \varphi_4 \rangle)]$

Example 1 Consider T which consists of $\varphi_1, \dots, \varphi_6$ as above in this section. Let $M = \{0.2/d\}$, and $PCLOSED = false$. After the initialization (line 16 of the algorithm), we have $NEWDEP = \{0.4/a, 0.2/d\}$ and $UPDATE = (\langle a, 0.4 \rangle, \langle d, 0.2 \rangle)$. Recall that during the update, values of $COUNT$ and $SKIP$ are changed. Namely, values of $COUNT$ may be decremented and values of $SKIP$ are changed to *true*. The latter update is represented by removing pointers from the T -structure. After the update of $\langle a, 0.4 \rangle$ and $\langle d, 0.2 \rangle$, the T -record $\langle 0, 0.6, \varphi_2 = \{0.4/a, 0.2/d\} \Rightarrow \{0.2/e\}$ of φ_2 is processed because we have $COUNT[\varphi_2] = 0$ (see the first item of the T -record). At this point, the algorithm is in the following state:

b : $[(1, \langle 2, 2, \varphi_6 \rangle), \langle 2, 1.1, \varphi_5 \rangle]$ $ADD = (\langle e, 0.2 \rangle)$
 c : $[(0.5, \langle 3, 1.3, \varphi_4 \rangle), (1, \langle 2, 2, \varphi_6 \rangle)]$ $NEWDEP = \{0.4/a, 0.2/d, 0.2/e\}$
 d : $[(0.4, \langle 3, 1.3, \varphi_4 \rangle)]$ $UPDATE = (\langle e, 0.2 \rangle)$
 e : $[(0.1, \langle 2, 1.1, \varphi_5 \rangle), (0.2, \langle 1, 0.4, \varphi_3 \rangle), (0.4, \langle 3, 1.3, \varphi_4 \rangle)]$

The corresponding T -structure is depicted in Fig. 4. As a further step of the computation, an update of $\langle e, 0.2 \rangle$ is performed and then the T -record $\langle 0, 0.4, \varphi_3 = \{0.2/d, 0.2/e\} \Rightarrow \{0.6/c, 0.5/d, 0.5/e\}$ of φ_3 is processed:

b : $[(1, \langle 2, 2, \varphi_6 \rangle), \langle 1, 1.1, \varphi_5 \rangle]$ $ADD = (\langle c, 0.6 \rangle, \langle d, 0.5 \rangle, \langle e, 0.5 \rangle)$
 c : $[(0.5, \langle 3, 1.3, \varphi_4 \rangle), (1, \langle 2, 2, \varphi_6 \rangle)]$ $NEWDEP = \{0.4/a, 0.6/c, 0.5/d, 0.5/e\}$
 d : $[(0.4, \langle 3, 1.3, \varphi_4 \rangle)]$ $UPDATE = (\langle c, 0.6 \rangle, \langle d, 0.5 \rangle, \langle e, 0.5 \rangle)$
 e : $[(0.4, \langle 3, 1.3, \varphi_4 \rangle)]$

Right after the update of $\langle c, 0.6 \rangle$, $\langle d, 0.5 \rangle$, and $\langle e, 0.5 \rangle$, the algorithm will process the T -record of φ_4 . After that, we have the following situation:

b : $[(1, \langle 2, 2, \varphi_6 \rangle), \langle 1, 1.1, \varphi_5 \rangle]$ $ADD = (\langle a, 0.8 \rangle, \langle b, 1 \rangle)$
 c : $[(1, \langle 2, 2, \varphi_6 \rangle)]$ $NEWDEP = \{0.8/a, b, 0.6/c, 0.5/d, 0.5/e\}$
 $UPDATE = (\langle a, 0.8 \rangle, \langle b, 1 \rangle)$

Then, $\langle a, 0.8 \rangle$ is updated. Notice that this update has no effect because the T -structure no longer contains attributes of the form $\langle a, x \rangle$ waiting for update (the

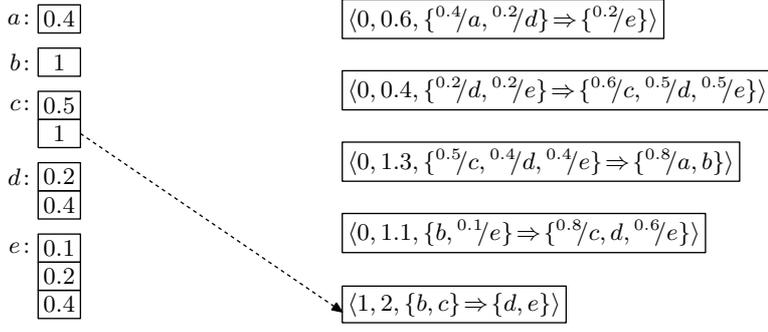


Figure 5: T -structure after the computation

a -list is empty). After the update of $\langle b, 1 \rangle$, the T -record $\langle 0, 1.1, \varphi_5 = \{b, 0.1/e\} \Rightarrow \{0.8/c, d, 0.6/e\}$ of φ_5 is processed. We arrive to:

$$\begin{aligned} c: [(1, \langle 1, 2, \varphi_6 \rangle)] \quad & ADD = (\langle c, 0.8 \rangle, \langle d, 1 \rangle, \langle e, 0.6 \rangle) \\ & NEWDEP = \{0.8/a, b, 0.8/c, d, 0.6/e\} \\ & UPDATE = (\langle c, 0.8 \rangle, \langle d, 1 \rangle, \langle e, 0.6 \rangle) \end{aligned}$$

The algorithm updates $\langle c, 0.8 \rangle$, $\langle d, 1 \rangle$, $\langle e, 0.6 \rangle$ however such updates are all without any effect because the d -list and e -list are already empty, and the c -list contains a single record with $1 \not\leq 0.8$ (see the condition at line 20 of the algorithm). Thus, the T -structure remains unchanged (see Fig. 5), $UPDATE$ is empty, and the procedure stops returning the value of $NEWDEP$ which is $\{0.8/a, b, 0.8/c, d, 0.6/e\}$.

Example 2 In this example we demonstrate the role of the *WAITLIST*. Let T be a set of FAIs which consists of

$$\begin{aligned} \psi_1: \{0.2/a\} \Rightarrow \{0.6/a, 0.3/c\}, & \quad \psi_3: \{0.6/a, 0.3/c\} \Rightarrow \{b\}, \\ \psi_2: \{0.3/c\} \Rightarrow \{0.2/b\}, & \quad \psi_4: \{0.6/a, b, 0.3/c\} \Rightarrow \{d\}. \end{aligned}$$

Moreover, we consider $M = \{0.3/a\}$ and $PCLOSED = true$. After the initialization (line 16), we have $NEWDEP = \{0.3/a\}$, $CARDND = 0.3$ (f_L is identity), $UPDATE = (\langle a, 0.3 \rangle)$, $WAITLIST = ()$, and the T -structure is the following:

$$\begin{aligned} a: [(0.2, \langle 1, 0.2, \psi_1 \rangle), (0.6, \langle 3, 1.9, \psi_4 \rangle), \langle 2, 0.9, \psi_3 \rangle)] \\ b: [(1, \langle 3, 1.9, \psi_4 \rangle)] \\ c: [(0.3, \langle 3, 1.9, \psi_4 \rangle), \langle 2, 0.9, \psi_3 \rangle, \langle 1, 0.3, \psi_2 \rangle)] \end{aligned}$$

The computation continues with the update of $\langle a, 0.3 \rangle$. During that, the T -record $\langle 1, 0.2, \psi_1 \rangle$ will be updated to $\langle 0, 0.2, \psi_1 \rangle$. Since $CARD[\psi_1] = 0.2 < 0.3 = CARDND$, the left-hand side of ψ_1 is strictly contained in $NEWDEP$, and the algorithm processes $\langle 0, 0.2, \psi_1 = \{0.2/a\} \Rightarrow \{0.6/a, 0.3/c\}$, i.e. we get to

$$\begin{aligned} a: [(0.6, \langle 3, 1.9, \psi_4 \rangle), \langle 2, 0.9, \psi_3 \rangle] \quad & ADD = (\langle a, 0.6 \rangle, \langle c, 0.3 \rangle) \\ b: [(1, \langle 3, 1.9, \psi_4 \rangle)] \quad & NEWDEP = \{0.6/a, 0.3/c\} \\ c: [(0.3, \langle 3, 1.9, \psi_4 \rangle), \langle 2, 0.9, \psi_3 \rangle, \langle 1, 0.3, \psi_2 \rangle] \quad & CARDND = 0.9 \\ & UPDATE = (\langle a, 0.6 \rangle, \langle c, 0.3 \rangle) \end{aligned}$$

After the update of $\langle a, 0.6 \rangle$, we have:

$$\begin{aligned} b: [(1, \langle 2, 1.9, \psi_4 \rangle)] \\ c: [(0.3, \langle 2, 1.9, \psi_4 \rangle), \langle 1, 0.9, \psi_3 \rangle, \langle 1, 0.3, \psi_2 \rangle)] \end{aligned}$$

	GDP		unemployment	
	high (gh)	low (gl)	high (uh)	low (ul)
Algeria	0.0	0.3	1.0	0.0
Austria	0.8	0.0	0.0	1.0
Belgium	0.7	0.0	0.7	0.0
Botswana	0.1	0.1	1.0	0.0
Cote d'Ivoire	0.0	0.8	0.8	0.0
Croatia	0.2	0.0	0.9	0.0
Denmark	0.8	0.0	0.1	0.8
Djibouti	0.0	0.9	1.0	0.0
Egypt	0.0	0.6	0.6	0.0
Estonia	0.3	0.0	0.5	0.1

Table 1: Data table with fuzzy attributes

$$\begin{aligned}
\{^{0.1}/ul\} &\Rightarrow \{^{0.3}/gh, ^{0.1}/ul\}, & \{^{0.1}/gh, ^{0.8}/uh\} &\Rightarrow \{^{0.1}/gh, ^{0.9}/uh\}, \\
\{^{0.2}/uh\} &\Rightarrow \{^{0.5}/uh\}, & \{^{0.1}/gh, ^{0.3}/gl, uh\} &\Rightarrow \{gh, gl, uh, ul\}, \\
\{uh\} &\Rightarrow \{^{0.1}/gl, uh\}, & \{^{0.2}/gh, ^{0.1}/gl, uh\} &\Rightarrow \{gh, gl, uh, ul\}, \\
\{^{0.1}/gl\} &\Rightarrow \{^{0.1}/gl, ^{0.6}/uh\}, & \{^{0.3}/gh, ^{0.2}/ul\} &\Rightarrow \{^{0.8}/gh, ^{0.8}/ul\}, \\
\{^{0.1}/gl, ^{0.7}/uh\} &\Rightarrow \{^{0.1}/gl, ^{0.8}/uh\}, & \{^{0.3}/gh, ^{0.7}/uh\} &\Rightarrow \{^{0.7}/gh, ^{0.7}/uh\}, \\
\{^{0.1}/gl, ^{0.9}/uh\} &\Rightarrow \{^{0.1}/gl, uh\}, & \{^{0.4}/gh\} &\Rightarrow \{^{0.7}/gh\}, \\
\{^{0.2}/gl, ^{0.6}/uh\} &\Rightarrow \{^{0.3}/gl, ^{0.6}/uh\}, & \{^{0.7}/gh, ^{0.1}/ul\} &\Rightarrow \{^{0.8}/gh, ^{0.8}/ul\}, \\
\{^{0.4}/gl, ^{0.6}/uh\} &\Rightarrow \{^{0.6}/gl, ^{0.6}/uh\}, & \{^{0.7}/gh, ^{0.5}/uh\} &\Rightarrow \{^{0.7}/gh, ^{0.7}/uh\}, \\
\{^{0.6}/gl, ^{0.8}/uh\} &\Rightarrow \{^{0.8}/gl, ^{0.8}/uh\}, & \{^{0.7}/gh, ^{0.9}/uh\} &\Rightarrow \{gh, gl, uh, ul\}, \\
\{^{0.7}/gl, ^{0.6}/uh\} &\Rightarrow \{^{0.8}/gl, ^{0.8}/uh\}, & \{^{0.8}/gh\} &\Rightarrow \{^{0.8}/gh, ^{0.8}/ul\}, \\
\{^{0.8}/gl, uh\} &\Rightarrow \{^{0.9}/gl, uh\}, & \{^{0.8}/gh, ^{0.9}/ul\} &\Rightarrow \{^{0.8}/gh, ul\}, \\
\{^{0.9}/gl, ^{0.8}/uh\} &\Rightarrow \{^{0.9}/gl, uh\}, & \{^{0.8}/gh, ^{0.1}/uh, ul\} &\Rightarrow \{gh, gl, uh, ul\}, \\
\{gl, uh\} &\Rightarrow \{gh, gl, uh, ul\}, & \{^{0.8}/gh, ^{0.7}/uh, ^{0.8}/ul\} &\Rightarrow \{gh, gl, uh, ul\}, \\
\{^{0.1}/gh, ^{0.6}/uh\} &\Rightarrow \{^{0.1}/gh, ^{0.7}/uh\}, & \{^{0.9}/gh, ^{0.8}/ul\} &\Rightarrow \{gh, gl, uh, ul\}.
\end{aligned}$$

Figure 6: Minimal basis of data table

Then, the algorithm continues with updating $\langle c, 0.3 \rangle$. The T -record $\langle 2, 1.9, \psi_4 \rangle$ is updated to $\langle 1, 1.9, \psi_4 \rangle$ and removed from the c -list. In the next step, the T -record $\langle 1, 0.9, \psi_3 \rangle$ is updated to $\langle 0, 0.9, \psi_3 \rangle$. At this point, we have $CARD[\psi_3] = 0.9 = CARDND$, i.e. we add fuzzy set $\{b\}$ of attributes (the right-hand side of ψ_3) to the $WAITLIST$. Finally, $\langle 1, 0.3, \psi_2 \rangle$ is updated to $\langle 0, 0.3, \psi_2 \rangle$ which yields the following situation: the T -structure consists of $b: [(1, \langle 1, 1.9, \psi_4 \rangle)]$, $ADD = (\langle b, 0.2 \rangle)$, $NEWDEP = \{^{0.6}/a, ^{0.2}/b, ^{0.3}/c\}$, $CARDND = 1.1$, and $UPDATE = (\langle b, 0.2 \rangle)$. Since ADD is nonempty, the algorithm continues with flushing the $WAITLIST$ (lines 28–35). After that, the new values are set to $NEWDEP = \{^{0.6}/a, b, ^{0.3}/c\}$, $CARDND = 1.9$, and $UPDATE = (\langle b, 0.2 \rangle, \langle b, 1 \rangle)$. The process continues with updating $\langle b, 0.2 \rangle$ (no effect) and $\langle b, 1 \rangle$. Here again, we are in a situation where $CARD[\psi_4] = 1.9 = CARDND$, i.e. $\{d\}$ is added to the $WAITLIST$, only this time, the computation ends because $UPDATE$ is empty, i.e. $\{d\}$ will not be added to $NEWDEP$. Thus, the resulting value being returned is $\{^{0.6}/a, b, ^{0.3}/c\}$.

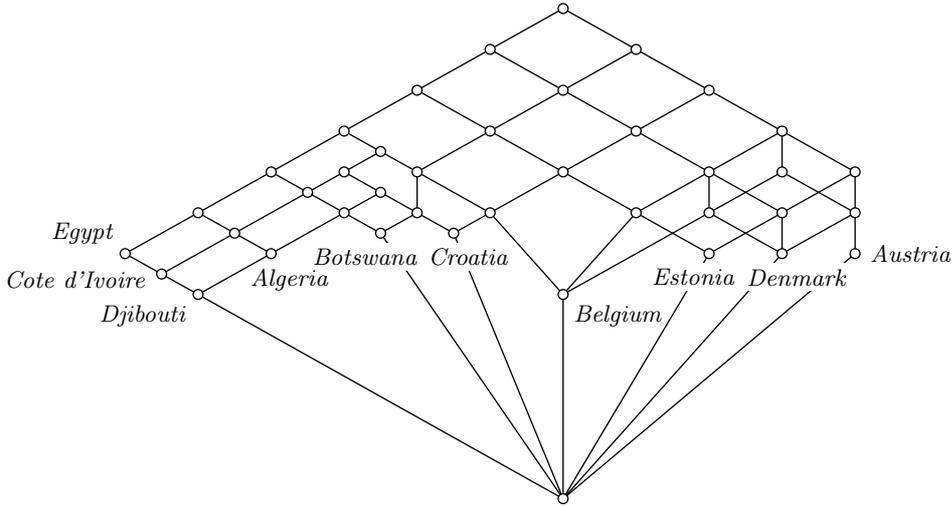


Figure 7: Concept lattice extracted from data table

Example 3 For illustration, we present a non-redundant (and minimal) basis which is computed by our algorithm in combination with extension of Ganter’s NEXTCLOSURE algorithm, see [2, 5, 14, 15, 16]. Our structure of truth degrees will be a linearly ordered residuated lattice \mathbf{L} with globalization such that $L = \{0, 0.1, 0.2, \dots, 0.9, 1\}$. Input data is represented by data table in Table 1. Objects (rows of data table) are selected countries and the attributes are “high GDP”, “low GDP”, “high unemployment”, and “low unemployment”. Table entries are degree from L indicating to what degrees countries have corresponding attributes. A non-redundant basis of the table is shown in Fig. 6. Recall that the basis is a set T of FAIs such that all FAIs which are true in the table are exactly the FAIs which are entailed from the basis. Also note that models of the basis correspond to formal concepts (particular conceptual clusters) which can be found in the data tables. Namely, fuzzy sets of attributes which fall under concepts (so-called intents) are exactly the models of the basis. The hierarchy of concepts found in Table 1 (and thus, the hierarchy of models of the basis) is depicted in Fig. 7. More details on model-theoretical properties of FAIs can be found in [5, 11].

Remark 8 Compared to the naive way to compute fixed points of operators cl_T and cl_{T^*} , which has been discussed in the beginning of Section 3, GLINCLOSURE is considerably faster. Preliminary tests of efficiency were presented in [21]. The algorithm was used in conjunction with graded extension of Ganter’s NextClosure algorithm to compute bases of data tables populated with entries generated according to selected distributions. The tests have shown that the increase of performance over CLOSURE (the naive algorithm) is especially remarkable if we work with dense data tables, i.e. data tables with small number of entries containing 0. For illustration, the graph in Fig. 8 shows performance of GLINCLOSURE for data tables with 15 attributes. The time is measured in seconds (logarithmic scale). The graph

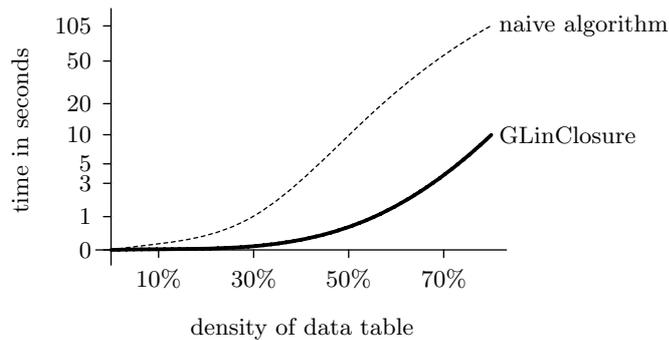


Figure 8: Performance of GLINCLOSURE compared to the naive algorithm

shows that on dense data sets with 15 attributes, GLINCLOSURE is approximately ten times faster. The efficiency of GLINCLOSURE and related algorithms will be the subject of further study.

5. Conclusions

We have shown an extended version of the LINCLOSURE algorithm, so-called GRADED LINCLOSURE (GLINCLOSURE). Our algorithm can be used in case of graded as well as binary attributes. Even for binary attributes, GLINCLOSURE is more versatile than the original LINCLOSURE (it can be used to compute systems of pseudo-intents) but it has the same asymptotic complexity $O(n)$. Future research will focus on further algorithms for formal concept analysis of data with fuzzy attributes.

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, M. Chlupova, V. Vychodil: “Implications from data with fuzzy attributes,” in *AISTA 2004 in Cooperation with the IEEE Computer Society Proceedings*, 2004, 5 pages, ISBN 2-9599776-8-8.
3. R. Belohlavek, T. Funiokova, V. Vychodil: “Fuzzy closure operators with truth stressers,” *Logic Journal of IGPL* **13**(5)(2005), 503–513.
4. R. Belohlavek, V. Vychodil: “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.
5. R. Belohlavek, V. Vychodil: “Fuzzy attribute logic: attribute implications, their

- validity, entailment, and non-redundant basis,” in Liu Y., Chen G., Ying M. (Eds.): *Fuzzy Logic, Soft Computing & Computational Intelligence: Eleventh International Fuzzy Systems Association World Congress* (Vol. I), 2005, pp. 622–627. Tsinghua University Press and Springer, ISBN 7–302–11377–7.
6. R. Belohlavek, V. Vychodil: “Fuzzy attribute implications: computing non-redundant bases using maximal independent sets”, in S. Zhang and R. Jarvis (Eds.): *AI 2005*, LNAI **3809**, pp. 1126–1129, 2005.
 7. R. Belohlavek, V. Vychodil: “Attribute implications in a fuzzy setting,” in Missaoui R., Schmid J. (Eds.): *ICFCA 2006*, LNAI **3874**, pp. 45–60, 2006.
 8. R. Belohlavek, V. Vychodil: “Functional dependencies of data tables over domains with similarity relations,” in *Proc. IICAI 2005*, pp. 2486–2504, ISBN 0–9727412–1–6.
 9. R. Belohlavek, V. Vychodil: “Data tables with similarity relations: functional dependencies, complete rules and non-redundant bases,” in Lee M. L., Tan K. L., Wuwongse V. (Eds.): *DASFAA 2006*, LNCS **3882**, pp. 644–658, 2006.
 10. R. Belohlavek, V. Vychodil: “Computing non-redundant bases of if-then rules from data tables with graded attributes,” in Zhang Y. Q., Lin T. Y. (Eds.): *Proc. IEEE-GrC 2006*, pp. 205–210, 2006.
 11. R. Belohlavek, V. Vychodil: “Properties of models of fuzzy attribute implications,” in *Proc. SCIS & ISIS 2006: Joint 3rd International Conference on Soft Computing and Intelligent Systems and 7th International Symposium on advanced Intelligent Systems*, pp. 291–296, Tokyo Institute of Technology, Japan Society for Fuzzy Theory and Intelligent Informatics, 2006, ISSN 1880–3741.
 12. R. Belohlavek, V. Vychodil: “Fuzzy attribute logic over complete residuated lattices,” *J. Exp. Theor. Artif. Intelligence* **18**(4)(2006), pp. 471–480.
 13. C. Carpineto, G. Romano: *Concept Data Analysis. Theory and Applications*. J. Wiley, 2004.
 14. B. Ganter: *Begriffe und Implikationen*, manuscript, 1998.
 15. B. Ganter: “Algorithmen zur formalen Begriffsanalyse,” in B. Ganter, R. Wille, K.E. Wolff (Hrsg.): *Beiträge zur Begriffsanalyse*. B. I. Wissenschaftsverlag, Mannheim, 1987, 241–254.
 16. B. Ganter, R. Wille: *Formal Concept Analysis. Mathematical Foundations*. Springer, Berlin, 1999.
 17. J. A. Goguen: “The logic of inexact concepts,” *Synthese* **18**(1968-9), 325–373.
 18. J.-L. Guigues, V. Duquenne: “Familles minimales d’implications informatives resultant d’un tableau de données binaires,” *Math. Sci. Humaines* **95**(1986), 5–18.
 19. P. Hájek: *Metamathematics of Fuzzy Logic*. Kluwer, Dordrecht, 1998.
 20. P. Hájek: “On very true,” *Fuzzy Sets and Systems* **124**(2001), 329–333.
 21. Z. Horák: “Exploring dependencies in vague data,” UP Olomouc (MSc. thesis, in Czech), 2006.
 22. G. J. Klir, B. Yuan: *Fuzzy Sets and Fuzzy Logic. Theory and Applications*. Prentice Hall, 1995.
 23. D. Maier: *The Theory of Relational Databases*. Computer Science Press, Rockville, 1983.
 24. J. Pavelka: “On fuzzy logic I, II, III,” *Z. Math. Logik Grundlagen Math.* **25**(1979), 45–52, 119–134, 447–464.
 25. S. Pollandt: *Fuzzy Begriffe*. Springer-Verlag, Berlin/Heidelberg, 1997.

26. G. Takeuti, S. Titani: "Globalization of intuitionistic set theory," *Annals of Pure and Applied Logic* **33**(1987), 195–211.