Fuzzy Attribute Implications: Computing Non-redundant Bases Using Maximal Independent Sets*

Radim Bělohlávek and Vilém Vychodil

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

Abstract. This note describes 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. We reduce the problem of computing systems of pseudo-intents to the problem of computing maximal independent sets in certain graphs. We outline theoretical foundations, the algorithm, and present demonstrating examples.

1 Introduction and Problem Setting

Fuzzy attribute implications are formulas of a form "if A then 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 [8,10]. 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 [4,7]. 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, there is a unique system of pseudo-intents and an algorithm for the computation of the corresponding non-redundant basis was presented in [4]. The present paper shows an algorithm for getting systems of pseudo-intents for general hedges. Due to the limited scope of this paper, we postpone all proofs to a forthcoming paper.

* 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.

S. Zhang and R. Jarvis (Eds.): AI 2005, LNAI 3809, pp. 1126–1129, 2005.
© Springer-Verlag Berlin Heidelberg 2005

2 Fuzzy Attribute Logic (Preliminaries)

In this section we survey notions of fuzzy attribute logic, for more details see [4,5,7]. We use residuated structures of truth degrees [2] and the usual notions of fuzzy logic and fuzzy sets. A residuated structure of truth degrees will be denoted by $\mathbf{L}$. Let Y be a finite set of attributes. A (*fuzzy*) *attribute implication* 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". A data table with fuzzy attributes [4,5,7] is a triplet $\mathcal{T} = \langle X, Y, I \rangle$ where X is a set of objects, Y is a finite set of attributes, and $I \in \mathbf{L}^{X \times Y}$ is a binary $\mathbf{L}$-relation (fuzzy 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. $\mathcal{T}$ can be seen as a table with rows and columns corresponding to objects $x \in X$ and attributes $y \in Y$, respectively, and table entries containing degrees $I(x, y)$. For fuzzy set $M \in \mathbf{L}^Y$ of attributes, we define a *degree* $||A \Rightarrow B||_M$ *to which* $A \Rightarrow B$ *is true in* M by $||A \Rightarrow B||_M = S(A, M)^* \rightarrow S(B, M)$, where $\rightarrow$ is residuated implication (in $\mathbf{L}$), * is a hedge, and $S(A, M)$ denotes degree of subsethood of A in M, for details see [2,4,5,7,11]. A row of a table $\mathcal{T}$ corresponding to an object $x \in X$ can be seen as a fuzzy set I_x of attributes to which each $y \in Y$ belongs to a degree $I_x(y) = I(x, y)$. Given $\mathcal{T} = \langle X, Y, I \rangle$, we define *degree* $||A \Rightarrow B||_{\langle X,Y,I \rangle}$ *to which* $A \Rightarrow B$ *is true in* (*each row of*) $\mathcal{T}$ by $||A \Rightarrow B||_{\langle X,Y,I \rangle} = \bigwedge_{x \in X} ||A \Rightarrow B||_{I_x}$. Let T be a set of fuzzy attribute implications. $M \in \mathbf{L}^Y$ 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 is denoted by $\mathrm{Mod}(T)$. A degree $||A \Rightarrow B||_T$ to which $A \Rightarrow B$ *semantically follows* from T is defined by $||A \Rightarrow B||_T = \bigwedge_{M \in \mathrm{Mod}(T)} ||A \Rightarrow B||_M$. 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, 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* $\mathcal{T} = \langle X, Y, I \rangle$). For $A \in \mathbf{L}^X$, $B \in \mathbf{L}^Y$, we define $A^{\uparrow} \in \mathbf{L}^Y$, $B^{\downarrow} \in \mathbf{L}^X$ by $A^{\uparrow}(y) = \bigwedge_{x \in X}\big(A(x)^* \rightarrow I(x, y)\big)$, $B^{\downarrow}(x) = \bigwedge_{y \in Y}\big(B(y) \rightarrow I(x, y)\big)$. Operators $^{\downarrow}, ^{\uparrow}$ form so-called Galois connection with hedge [6], composed operator $^{\downarrow\uparrow}$ is a fuzzy closure operator. The (lattice-ordered) structure of all fixed points of $^{\downarrow\uparrow}$ is called a fuzzy concept lattice (induced by $\mathcal{T}$), see [2,3,6]. Given $\mathcal{T} = \langle X, Y, I \rangle$, 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: $P \in \mathcal{P}$ if and only if $P \neq P^{\downarrow\uparrow}$ and, for each $Q \in \mathcal{P}$ such that $Q \neq P$, $||Q \Rightarrow Q^{\downarrow\uparrow}||_P = 1$. The following property of systems of pseudo-intents was proved is [7]: if $\mathcal{P}$ is a system of pseudo-intents of $\mathcal{T}$ then $T = \{P \Rightarrow P^{\downarrow\uparrow} \mid P \in \mathcal{P}\}$ is a non-redundant basis of $\mathcal{T}$. Therefore we focus on computing of systems of pseudo-intents.

3 Computing Non-redundant Bases

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

$$V = \{P \in \mathbf{L}^Y \mid P \neq P^{\downarrow\uparrow}\}. \tag{1}$$

If $V \neq \emptyset$, we define a binary relation E on V by

$$E = \{\langle P, Q\rangle \in V \mid P \neq Q \text{ and } ||Q \Rightarrow Q^{\downarrow\uparrow}||_P \neq 1\}. \quad (2)$$

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

Theorem. *Let $\mathbf{L}$ be a finite residuated structure of truth degrees, * be any (truth-stressing) hedge, $\mathcal{T} = \langle X, Y, I\rangle$ be a data table with fuzzy attributes (with truth degrees in $\mathbf{L}$), $\mathcal{P} \subseteq \mathbf{L}^Y$, V and E be defined by* (1) *and* (2)*, respectively. Then the following statements are equivalent.*
(i) *$\mathcal{P}$ is a system of pseudo-intents;*
(ii) *$V - \mathcal{P} = \text{Pred}(\mathcal{P})$;*
(iii) *$\mathcal{P}$ is a maximal independent set in $\mathbf{G}$ such that $V - \mathcal{P} = \text{Pred}(\mathcal{P})$.* □

The Theorem gives 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 $V - \mathcal{P} = \text{Pred}(\mathcal{P})$. If * is globalization on finite $\mathbf{L}$, for each $\mathcal{T}$ there is exactly one system of pseudo-intents, see [4,7], which can also be found by the graph procedure but we can take 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 (details are postponed to the full version of the paper).

4 Example 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 Fig. 1 (left). The set X of object consists of objects "Mercury", "Venus", …; Y contains four attributes: size of a planet (small / large), distance from the sun (far / near). If * (hedge) is defined so that $1^* = 1$ and $a^* = 0$ for all $a < 1$ (so-called globalization), there is a unique system of pseudo-intents, see [4]. If, for each $a \in L$, $a^* = a$ (i.e., * is identity), we obtain two distinct systems of pseudo-intents (denote the systems by $\mathcal{P}_1$ and $\mathcal{P}_2$), both consist of four elements. Fig. 1 (right) contains the incidence matrix of relation $E \subseteq V \times V$ defined by (2). 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 $\langle P, Q\rangle \in E$, "○" ("×") indicates that $Q \in \mathcal{P}_1$ ($Q \in \mathcal{P}_2$) and $P \in \text{Pred}(Q)$. Non-redundant bases T_1 and T_2 given by $\mathcal{P}_1$ and $\mathcal{P}_2$ are the following:

$$\begin{aligned} T_1 = \{&\{n\} \Rightarrow \{{}^{0.5}\!/s, n\}, \{f, {}^{0.5}\!/n\} \Rightarrow \{{}^{0.5}\!/l, f, {}^{0.5}\!/n\},\\ &\{l\} \Rightarrow \{l, f, {}^{0.5}\!/n\}, \{s, {}^{0.5}\!/l, {}^{0.5}\!/f\} \Rightarrow \{s, {}^{0.5}\!/l, {}^{0.5}\!/f, {}^{0.5}\!/n\}\},\\ T_2 = \{&\{n\} \Rightarrow \{{}^{0.5}\!/s, n\}, \{f, {}^{0.5}\!/n\} \Rightarrow \{{}^{0.5}\!/l, f, {}^{0.5}\!/n\},\\ &\{l, {}^{0.5}\!/f\} \Rightarrow \{l, f, {}^{0.5}\!/n\}, \{s, {}^{0.5}\!/l, {}^{0.5}\!/f\} \Rightarrow \{s, {}^{0.5}\!/l, {}^{0.5}\!/f, {}^{0.5}\!/n\}\}. \end{aligned}$$

If $\mathbf{L}$ is a three-element Gödel chain with identity, we get two distinct systems of pseudo-intents with various sizes. Experiments with randomly generated data have shown 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.

	size		distance	
	small (s)	large (l)	far (f)	near (n)
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

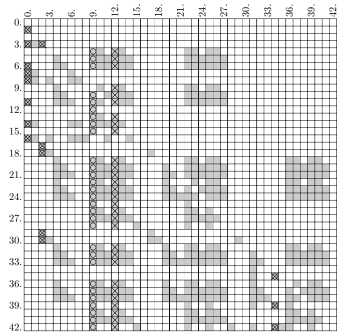

Fig. 1. Data table with fuzzy attributes and the corresponding V and E

References

1. Bělohlávek R.: Similarity relations in concept lattices. *J. Logic Comput.* 10(6):823–845, 2000.
2. Bělohlávek R.: *Fuzzy Relational Systems: Foundations and Principles.* Kluwer, Academic/Plenum Publishers, New York, 2002.
3. Bělohlávek R.: Concept lattices and order in fuzzy logic. *Ann. Pure Appl. Logic* **128**(2004), 277–298.
4. 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.
5. 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, pp. 1050–1055, ISBN 0–7803–9158–6.
6. Bělohlávek R., Vychodil V.: Reducing attribute implications from data tables with fuzzy attributes to tables with binary attributes. *Proceedings of the 8th Joint Conference on Information Sciences*, 2005, pp. 78–81, ISBN 0–9707890–3–3.
7. Bělohlávek R., Vychodil V.: Fuzzy attribute logic: attribute implications, their validity, entailment, and non-redundant basis. In: Y. Liu, G. Chen, M. Ying (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.
8. Ganter B., Wille R.: *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin, 1999.
9. Gerla G.: *Fuzzy Logic. Mathematical Tools for Approximate Reasoning.* Kluwer, Dordrecht, 2001.
10. Guigues J.-L., Duquenne V.: Familles minimales d'implications informatives resultant d'un tableau de données binaires. *Math. Sci. Humaines* **95**(1986), 5–18.
11. Hájek P.: *Metamathematics of Fuzzy Logic.* Kluwer, Dordrecht, 1998.
12. Hájek P.: On very true. *Fuzzy Sets and Systems* **124**(2001), 329–333.
13. Klir G. J., Yuan B.: *Fuzzy Sets and Fuzzy Logic. Theory and Applications.* Prentice Hall, 1995.
14. Takeuti G., Titani S.: Globalization of intuitionistic set theory. *Annals of Pure and Applied Logic* **33**(1987), 195–211.