Information Sciences 351 (2016) 90–100

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Computing sets of graded attribute implications with witnessed non-redundancy

Vilem Vychodil*

Department of Computer Science, Palacky University, Olomouc 17. listopadu 12, CZ-77146 Olomouc, Czech Republic

ARTICLE INFO

Article history:
Received 9 November 2015
Revised 8 February 2016
Accepted 4 March 2016
Available online 9 March 2016

Keywords:
Redundancy
Attribute implication
Residuated lattice
Fuzzy logic
Pseudo intent

ABSTRACT

In this paper we extend our previous results on sets of graded attribute implications with witnessed non-redundancy. We assume finite residuated lattices as structures of truth degrees and use arbitrary idempotent truth-stressing linguistic hedges as parameters which influence the semantics of graded attribute implications. In this setting, we introduce algorithm which transforms any set of graded attribute implications into an equivalent non-redundant set of graded attribute implications with saturated consequents whose non-redundancy is witnessed by antecedents of the formulas. As a consequence, we solve the open problem regarding the existence of general systems of pseudo-intents which appear in formal concept analysis of object-attribute data with graded attributes and linguistic hedges. Furthermore, we show a polynomial-time procedure for determining bases given by general systems of pseudo-intents from sets of graded attribute implications which are complete in data.

1. Introduction

In this paper, we investigate properties of sets of graded attribute implications and extend the results presented in our recent paper [43]. The graded attribute implications, sometimes called fuzzy attribute implications [3], are rules describing if-then dependencies in data with graded attributes. The rules have been proposed and investigated from the point of view of formal concept analysis (shortly, FCA [24]) with linguistic hedges [15]. One of the basic problems in FCA is to extract, given a formal context, a set of attribute implications which is non-redundant and conveys the information about exactly all attribute implications which hold in the given formal context – such sets are called (non-redundant) bases of formal contexts. One of the most profound approaches of determining bases exploits the notion of a pseudo-intent which originated in [28] and has been later utilized, e.g., in [23]. The bases given by pseudo-intents are not only non-redundant but in addition minimal in terms of their cardinality. In our paper, we deal with a general notion of a system of pseudo-intents which appears in the generalization of FCA which includes graded attributes and uses linguistic hedges to reduce the size of concept lattices [8]. By a graded attribute we mean an attribute (property/feature) which may be satisfied (present) to degrees instead of just satisfied/not satisfied (present/not present) as in the ordinary setting. In the past, there have been many approaches to extensions of the traditional concept analysis which accommodate graded attributes [2,34,37,38] and related phenomena. Most of the approaches are focused solely on the structure of concept lattices with little or no attention paid to if-then rules. The exceptions seem the be the early works by Pollandt [38] and the results made in the framework

* Tel.: +420 585634705; fax: +420 585411643.
E-mail address: vychodil@binghamton.edu

http://dx.doi.org/10.1016/j.ins.2016.03.004

of FCA with linguistic hedges, see [16] for a survey. In [38], the author proposes generalized pseudo-intents which ensure that the constructed sets of formulas are complete in data, i.e., convey the information about exactly all if-then rules which hold in the data, but are redundant in general. Using a more general setting, [3,7] show that there is a general notion of a system of pseudo-intents which ensures both the completeness and non-redundancy. Unfortunately, the definition in [3] is not constructive and so far the procedure to find such systems was reduced to finding particular maximal independent sets of vertices in large graphs [6,10]. In addition, it has been shown that the existence and uniqueness of systems of pseudo-intents is not ensured in the general setting. Indeed, it follows that the properties of the underlying structures of truth degrees, which together with linguistic hedges determine the semantics of graded attribute implications, substantially affect the properties of such systems. In case of infinite structures of truth degrees, it is known that general systems of pseudo-intents may not exist [16]. The existence in case of finite structures was listed as one of the open problems in [35]. Our paper brings a positive answer to this question and shows that, among other results, that general systems of pseudo-intents can be determined in a polynomial time from any complete set of graded attribute implications. The result is based on some of our recent observations made in [43] where we have put in correspondence bases given by systems of pseudo-intents and non-redundant sets of graded attribute implications with saturated consequents where the non-redundancy of each formula is witnessed by its antecedent.

Detailed description of the problem and the results requires precise introduction of the utilized notions. Therefore, we postpone it to Section 3 after presenting the preliminaries in Section 2. In Section 3, we include the algorithm and comment on its immediate consequences. The soundness of the algorithm is proved in Section 4 which also contains additional remarks and examples. Finally, we present conclusions in Section 5.

2. Preliminaries

In this section, we present the basic notions related to the structures of truth degrees which are used in our paper and recall basic notions of graded attribute implications. We limit ourselves just to the notions which are utilized in this paper. Interested readers can find more details in [16]. Readers familiar with [43] can skip this section and go directly to Section 3.

A residuated lattice [2,22] is an algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$ where $\langle L, \wedge, \vee, 0, 1\rangle$ is a bounded lattice with 0 and 1 being the least and the greatest elements of L, respectively, $\langle L, \otimes, 1\rangle$ is a commutative monoid (i.e., $\otimes$ is commutative, associative, and 1 is neutral with respect to $\otimes$), and $\otimes$ and $\rightarrow$ satisfy the so-called adjointness property: for all $a, b, c \in L$, we have that $a \otimes b \leq c$ iff $a \leq b \rightarrow c$. Further in the paper, $\mathbf{L}$ always stands for a residuated lattice of the form $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$. $\mathbf{L}$ is a complete residuated lattice whenever $\langle L, \wedge, \vee, 0, 1\rangle$ is a complete lattice, (i.e., infima and suprema exist for arbitrary subsets of L). If L is finite, then $\mathbf{L}$ is trivially complete. Examples of (complete) residuated lattices include popular structures defined on the real unit interval using left-continuous triangular norms [20,32] and their finite substructures. The structures are utilized in mathematical fuzzy logics [18,26,27,29] and their applications [33] as structures of truth degrees with $\otimes$ and $\rightarrow$ used as truth functions of “fuzzy conjunction” and “fuzzy implication”, respectively.

As usual, a map A: $Y \rightarrow L$ is called an $\mathbf{L}$-set A in Y (or an $\mathbf{L}$-fuzzy set [25]); R: $X \times Y \rightarrow L$ is called a binary $\mathbf{L}$-relation between X and Y, $R(x, y) \in L$ is interpreted as the degree to which $x \in X$ and $y \in Y$ are related by R. The collection of all $\mathbf{L}$-sets in Y is denoted by L^Y. Operations with $\mathbf{L}$-sets are defined componentwise using operations in $\mathbf{L}$. For instance, the union $A \cup B$ of $\mathbf{L}$-sets $A \in L^Y$ and $B \in L^Y$ is an $\mathbf{L}$-set in Y such that $(A \cup B)(y) = A(y) \vee B(y)$; analogously for $\cap$ and $\wedge$. If $a \in L$ and $A \in L^Y$ then $a \otimes A$, called the a-multiple of A, is an $\mathbf{L}$-set in Y defined by $(a \otimes A)(y) = a \otimes A(y)$ for all $y \in Y$. For $A, B \in L^Y$, we define the degree $S(A, B)$ to which A is a subset of B by

$$S(A, B) = \textstyle\bigwedge_{y \in Y} \big(A(y) \rightarrow B(y)\big) \tag{1}$$

provided that the infimum of $\{A(y) \rightarrow B(y); y \in Y\} \subseteq L$ exists – this condition is satisfied, e.g., if $\mathbf{L}$ is complete or if Y is finite. Note that S given by (1) can be understood as a binary $\mathbf{L}$-relation on $\mathbf{L}$-sets, i.e., for a fixed Y, it is a map of the form S: $L^Y \times L^Y \rightarrow L$. It is easily seen that $S(A, B) = 1$ iff $A(y) \leq B(y)$ for all $y \in Y$ in which case we write $A \subseteq B$ and say that A is a full subset of B.

Remark 1. Let us note that the notion of a graded subsethood (1) defined using the residuated implication has been proposed by Goguen [25,26] and plays an important role in the interpretation of the if-then rules we consider in this paper. This corresponds with the fact that the usual inclusion of sets is used to define the interpretation of the classic attribute implications. Indeed, if Y is a non-empty set of attributes (symbolic names), any formula of the form $A \Rightarrow B$ where $A, B \subseteq Y$ is called an attribute implication [24]. Moreover, it is considered true given $M \subseteq Y$, written $M \models A \Rightarrow B$, whenever $A \subseteq M$ implies $B \subseteq M$. If we depart from the classic setting to the graded setting and replace the classic sets by $\mathbf{L}$-sets $A, B, M \in L^Y$, there are several possible ways how to define the notion of “$A \Rightarrow B$ being true in M” which all collapse to the ordinary notion when $\mathbf{L}$ is the two-element Boolean algebra. As it is described in detail in [16], two borderline (and both interesting) cases can be based on the graded and the full inclusion of $\mathbf{L}$-sets.

The framework we use in our paper enables us to reason with several different interpretations of inclusion of $\mathbf{L}$-sets, and thus several different ways of understanding the interpretation of data dependencies, using a single formalism which is based on additional parameterization of the semantics of the rules. Namely, we use the approach based on linguistic hedges [16]. In a more detail, given a non-empty and finite set Y of attributes and a complete residuated lattice $\mathbf{L}$, a graded attribute implication in Y is an expression $A \Rightarrow B$ where $A, B \in L^Y$; A is called the antecedent of $A \Rightarrow B$, B is called the