Information Sciences 294 (2015) 478–488

Contents lists available at ScienceDirect

Information Sciences

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

On minimal sets of graded attribute implications

Vilem Vychodil *

Dept. Computer Science, Palacky University, Olomouc, Czech Republic

ARTICLE INFO

Article history:
Received 9 June 2014
Received in revised form 22 August 2014
Accepted 28 September 2014
Available online 7 October 2014

Keywords:
Attribute implication
Fuzzy logic
Redundancy
Minimality
Residuated lattice

ABSTRACT

We explore the structure of non-redundant and minimal sets consisting of graded if-then rules. The rules serve as graded attribute implications in object-attribute incidence data and as similarity-based functional dependencies in a similarity-based generalization of the relational model of data. Based on our observations, we derive a polynomial-time algorithm which transforms a given finite set of rules into an equivalent one which has the least size in terms of the number of rules.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Reasoning with various types of if-then rules is crucial in many disciplines ranging from theoretical computer science to applications. Among the most widely used rules are those taking from of implications between conjunctions of attributes. Such rules are utilized in database systems (as functional dependencies or inclusion dependencies [23]), logic programming (as particular definite clauses representing programs [22]), and data mining (as attribute implications [14] or association rules [1,33]). One of the most important problems regarding the rules is to find for a given set T of rules a set of rules which is equivalent to T and minimal in terms of its size. In relational database theory [23], the problem is referred to as finding minimal covers of T.

In this paper, we deal with the problem of finding minimal and equivalent sets of rules for general rules describing dependencies between *graded attributes*. That is, instead of the classic rules which are often considered as implications

$$\{y_1,\ldots,y_m\} \Rightarrow \{z_1,\ldots,z_n\} \tag{1}$$

between sets of attributes, describing presence/absence of attributes, we deal with rules where the presence/absence of attributes is expressed to degrees. That is, the rules in question can be written as

$$\{a_1/y_1,\ldots,a_m/y_m\} \Rightarrow \{b_1/z_1,\ldots,b_n/z_n\} \tag{2}$$

and understood as rules saying that "if y_1 is present at least to degree a_1 and $\cdots$ and y_m is present at least to degree a_m, then z_1 is present at least to degree b_1 and $\cdots$ and z_n is present at least to degree b_n." We assume that the degrees appearing in (2) come from a structure of truth degrees which is more general than the two-element Boolean algebra and allows for *intermediate degrees of truth*. In particular, we use complete residuated lattices [13] with linguistic hedges [12,19,29] for the

* Tel.: +420 585 634 705; fax: +420 585 411 643.
E-mail address: vychodil@binghamton.edu

http://dx.doi.org/10.1016/j.ins.2014.09.059
0020-0255/© 2014 Elsevier Inc. All rights reserved.

job. In our setting, (2) can be seen as generalization of (1) if all the degrees $a_1, \ldots, b_1, \ldots$ are equal to 1 (as usual, 1 denotes the classical truth value of "full truth").

Our previous results on rules of the form (2) include a fixed point characterization of a semantic entailment, Armstrong-style [2] axiomatizations in the ordinary style and the graded style (also known as Pavelka-style completeness, see [24–26]), results on generating non-redundant bases from data, and two kinds of semantics of the rules: (i) a *database semantics* which is based on evaluating the rules in ranked data tables over domains with similarities [5], and (ii) an *incidence data semantics* which is based on evaluating the rules in object-attribute data tables with graded attributes [4,7] which are known as *formal contexts* in formal concept analysis [14]. Analogously as for the ordinary rules, one can show that both (i) and (ii) yield the same notion of semantic entailment which simplifies further considerations, e.g., a single axiomatization of the semantic entailment works for both the database and incidence data semantics of the rules. A survey of recent results regarding the rules can be found in [8].

In this paper, we consider rules like (2) and explore the structure of non-redundant and minimal sets of rules of this type. We show an if-and-only-if criterion of minimality and a polynomial-time procedure which, given T, transforms T into an equivalent and minimal set of graded rules. Let us note that the previous results regarding minimality of sets of graded rules [8] were focused exclusively on sets of rules generated from data. That is, the input for such instance-based approaches is not a set T of rules. Instead, the input is assumed to be a structure (e.g., a formal context with graded attributes or a database table over domains with similarities) and the goal is to find a minimal set T of rules which entails exactly all the rules true in the structure. One particular example is an algorithm for generating graded counterparts to Guigues–Duquenne bases [17] described in [8]. In contrast, the problem studied in this paper is different. We assume that a set T of rules is already given (e.g., inferred from data or proposed by an expert) but it may not be minimal. Therefore, it is interesting to find a minimal set of rules which conveys the same information. Unlike the instance-based methods which belong to hard problems [10] even for the classic (non-graded) rules, the minimization method presented in this paper is polynomial and therefore tractable.

The present paper is organized as follows. Section 2 presents preliminaries from structures of degrees and graded if-then rules. Section 3 contains the new results.

2. Preliminaries

In this section, we present basic notions from structures of truth degrees and graded attribute implications which formalize rules like (2). We only present the notions and results which are sufficient to follow the results in Section 3. Interested readers may find more results in [4,8,13,15,18,20].

A (complete) residuated lattice [4,13] is an algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$ where $\langle L, \wedge, \vee, 0, 1\rangle$ is a (complete) lattice, $\langle L, \otimes, 1\rangle$ is a commutative monoid, and $\otimes$ (multiplication, a truth function of "fuzzy conjunction") and $\rightarrow$ (residuum, a truth function of "fuzzy implication") satisfy the adjointness property: $a \otimes b \leqslant c$ iff $a \leqslant b \rightarrow c$ $(a, b, c \in L)$. Examples of complete residuated lattices include structures on the real unit interval given by left-continuous t-norms [11,18] as well as finite structures of degrees.

If $U \neq \emptyset$, we can consider the direct power $\mathbf{L}^U = \langle L^U, \cap, \cup, \otimes, \rightarrow, {}^*, \emptyset_U, 1_U\rangle$ of $\mathbf{L}$. Each $A \in L^U$ is called an $\mathbf{L}$-set (s fuzzy set) A in universe U. That is, $A \in L^U$ is a map $A : U \rightarrow L, A(u)$ being interpreted as "the degree to which u belongs to A". Operations $\cap, \cup, \otimes, \ldots$ in $\mathbf{L}^U$ represent operations with $\mathbf{L}$-sets which are induced by the corresponding operations $\wedge, \vee, \otimes, \ldots$ in $\mathbf{L}$. Hence, e.g., $(A \cup B)(u) = A(u) \vee B(u)$ for each $u \in U$. Note that for the lattice order $\subseteq$ in $\mathbf{L}^U$ being induced by $\leqslant$, we have $A \subseteq B$ iff, for each $u \in U$, $A(u) \leqslant B(u)$. Therefore, $A \subseteq B$ denotes "full containment" of A in B. If $U = \{u_1, \ldots, u_n\}$ (U is finite), we adopt the usual conventions for writing $\mathbf{L}$-sets $A \in L^U$ as $\{a_1/u_1, \ldots, a_n/u_n\}$ meaning that $A(u_i) = a_i$ $(i = 1, \ldots, n)$. Furthermore, in the notation we omit a_i/u_i if $a_i = 0$ and write u_i if $a_i = 1$.

Let Y be a finite non-empty set of attributes (i.e., symbolic names). A *graded attribute implication* in Y is an expression $A \Rightarrow B$, where $A, B \in L^Y$. In our paper, graded attribute implications are regarded as formulas representing rules like (2). The interpretation of graded attribute implications is based on the notion of a *graded subsethood of* $\mathbf{L}$*-sets* in a similar way as the interpretation of the ordinary attribute implications [14] is based on the ordinary subsethood. In a more detail, for any $A, M \in L^Y$, we define a degree $S(A, M) \in L$ of subsethood of A in M by

$$S(A, M) = \bigwedge\nolimits_{y \in Y}(A(y) \rightarrow M(y)). \tag{3}$$

Clearly, $A \subseteq M$ (i.e., A is fully contained in M) iff $S(A, M) = 1$. For any $A, B, M \in L^Y$, we may put

$$||A \Rightarrow B||_M = \begin{cases} S(B, M), & \text{if } A \subseteq M, \\ 1, & \text{otherwise}, \end{cases} \tag{4}$$

and call $||A \Rightarrow B||_M$ a degree to which $A \Rightarrow B$ is true in M. Therefore, if M is regarded as an $\mathbf{L}$-set of attributes of an object with each $M(y)$ interpreted as the degree to which the object has attribute y, then $||A \Rightarrow B||_M$ is a degree to which the following statement is true: "If the object has all the attributes from A, then it has all the attributes from B". Interestingly, (4) is not the only possible (and reasonable) interpretation of $A \Rightarrow B$ in M. In fact, our approach in [8] is more general in that it defines $||A \Rightarrow B||^*_M$ by

Download English Version:

https://daneshyari.com/en/article/392262

Download Persian Version:

https://daneshyari.com/article/392262

Daneshyari.com