International Journal of Approximate Reasoning 70 (2016) 51–67

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Automated prover for attribute dependencies in data with grades

Radim Belohlavek[a], Pablo Cordero[b], Manuel Enciso[b], Ángel Mora[b], Vilem Vychodil[a]

[a] Dept. Computer Science, Palacky University, Olomouc, Czech Republic
[b] Universidad de Málaga, Spain

ARTICLE INFO

Article history:
Received 19 May 2015
Received in revised form 20 October 2015
Accepted 8 December 2015
Available online 15 December 2015

Keywords:
Attribute implication
Functional dependency
Fuzzy logic
Ordinal data
Ordinary and graded completeness
Similarity

ABSTRACT

We present a new axiomatization of logic for dependencies in data with grades, which includes ordinal data and data over domains with similarity relations, and an efficient reasoning method that is based on the axiomatization. The logic has its ordinary-style completeness characterizing the ordinary, bivalent entailment as well as the graded-style completeness characterizing the general, possibly intermediate degrees of entailment. A core of the method is a new inference rule, called the rule of simplification, from which we derive convenient equivalences that allow us to simplify sets of dependencies while retaining semantic closure. The method makes it possible to compute a closure of a given collection of attributes with respect to a collection of dependencies, decide whether a given dependency is entailed by a given collection of dependencies, and more generally, compute the degree to which the dependency is entailed by a collection of dependencies. We also present an experimental evaluation of the presented method.

1. Introduction

We present a complete axiomatization of a logic for dependencies in data with grades and an efficient reasoning method based on this axiomatization. The dependencies are expressed by formulas of the form

$$A \Rightarrow B, \tag{1}$$

such as

$$\{0.2/y_1, y_2\} \Rightarrow \{0.8/y_3\}. \tag{2}$$

Formulas of the form (2) have two different interpretations whose entailment relations coincide. First, an interpretation given by object-attribute data with grades in which (2) means: every object that has attribute y_1 to degree at least 0.2 and attribute y_2 to degree 1 (i.e. fully possesses y_2) has attribute y_3 to degree at least 0.8. Second, an interpretation given by ranked tables over domains with similarities—a particular extension of Codd's model of relational data—in which (2) means: every two tuples that are similar on attribute y_1 to degree at least 0.2 and are equal on attribute y_2 are similar on attribute y_3 to degree at least 0.8. We assume that the degrees form a partially ordered set equipped with particular

E-mail addresses: radim.belohlavek@acm.org (R. Belohlavek), pcordero@uma.es (P. Cordero), enciso@uma.es (M. Enciso), amora@ctima.uma.es (Á. Mora), vychodil@acm.org (V. Vychodil).

http://dx.doi.org/10.1016/j.ijar.2015.12.007
0888-613X/

aggregation operations. If 0 and 1 are the only degrees, the first interpretation coincides with the well-known attribute dependencies in binary data saying that presence of certain attributes implies presence of other attributes, and the second one with functional dependencies in the ordinary Codd's model.

The logic based on the presented axiomatization obeys two types of completeness. The ordinary-style completeness says that $A \Rightarrow B$ semantically follows from a set T of dependencies if and only if $A \Rightarrow B$ is provable from T. The graded-style completeness characterizes the possibly intermediate degrees of entailment in that it says that the degree to which $A \Rightarrow B$ semantically follows from T equals the appropriately defined degree to which $A \Rightarrow B$ is provable from T, leaving classic entailment and non-entailment particular cases. The logic enables a new method of automated reasoning whose efficiency derives from a new rule, called a simplification equivalence, which makes it possible to replace theories by equivalent but simpler ones. The algorithm we present computes a closure of a given collection of attributes with respect to a given collection of formulas. A simple modification of the algorithm results in a procedure that decides entailment, i.e. decides whether $A \Rightarrow B$ follows from T, and more generally, computes the degree to which $A \Rightarrow B$ follows from T. The experimental evaluation we present demonstrates that the proposed method is computationally efficient and outperforms the previously proposed method based on the classic Closure algorithm [20].

2. Preliminary notions and results

In this section, we present preliminaries from complete residuated lattices and fuzzy attribute logic. Details can be found in [3,13–16].

2.1. Complete residuated lattices and related structures

We use complete residuated lattices as the structures of degrees. A complete residuated lattice is an algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$ such that $\langle L, \wedge, \vee, 0, 1\rangle$ is a complete lattice with 0 and 1 being the least and greatest element, respectively; $\langle L, \otimes, 1\rangle$ is a commutative monoid (i.e. $\otimes$ is commutative, associative, and $a \otimes 1 = 1 \otimes a = a$ for each $a \in L$); $\otimes$ and $\rightarrow$ satisfy the following adjointness property:

$$a \otimes b \leq c \quad \text{iff} \quad a \leq b \rightarrow c \tag{3}$$

for any $a, b, c \in L$. As usual in the context of fuzzy logics in narrow sense, we interpret elements in L as degrees (of truth) with the following comparative meaning: if $a = \|\varphi\|_e$ and $b = \|\psi\|_e$ are degrees from L assigned to formulas φ and ψ by evaluation e and if $a \leq b$, then φ is less true than ψ under e. Operations $\otimes$ and $\rightarrow$ (residuum) represent truth functions of logical connectives "fuzzy conjunction" and "fuzzy implication". Note that from (3) it follows that $a \leq b$ iff $a \rightarrow b = 1$.

Remark 1. Complete residuated lattices include infinite as well as finite structures. For instance, a large family of structures defined on the real unit interval with its natural ordering with $\otimes$ being left-continuous t-norms and $\rightarrow$ being the corresponding residua [3,16,19]. An important family of finite and linearly ordered complete residuated lattices results by considering finite substructures on t-norm based complete residuated lattices on the real unit interval. In particular, for $L = \{0, 1\}$, $\mathbf{L}$ may be identified with the two-element Boolean algebra of classical logic. Namely, $\wedge$ and $\vee$ then become the truth functions of classical conjunction and disjunction, $\otimes$ coincides with $\wedge$, and $\rightarrow$ becomes the truth function of classical implication.

We equip complete residuated lattices with an additional unary connective: an idempotent truth-stressing hedge (shortly, a hedge) on a complete residuated lattice $\mathbf{L}$ is a map ${}^*\colon L \rightarrow L$ satisfying the following conditions: (i) $1^* = 1$, (ii) $a^* \leq a$, (iii) $(a \rightarrow b)^* \leq a^* \rightarrow b^*$, (iv) $a^{**} = a^*$ for each $a, b \in L$. Truth-stressing hedges were investigated from the point of view of fuzzy logic in narrow sense by Hájek [17], see also a recent general approach in [11,12]. In fuzzy logics, truth-stressing hedges serve as truth functions for unary connectives like "very true". For instance, Hájek [17] introduces a unary connective "vt" and formulas of the form $vt\varphi$ which read "φ is very true". Such formulas and evaluated (under e) so that $\|vt\varphi\|_e = (\|\varphi\|_e)^*$ with * being a truth-stressing hedge. Two important boundary cases of hedges are identity (i.e. $x^* = x$ for all $x \in L$) and so-called globalization [23] (i.e. $1^* = 1$ and $x^* = 0$ for all $1 \neq x \in L$).

Considering $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, 0, 1\rangle$ as the structure of degrees, we introduce the basic notions of fuzzy relational systems. An $\mathbf{L}$-set (a fuzzy set with degrees in $\mathbf{L}$) A in universe Y is any map $A\colon Y \rightarrow L$, $A(y)$ being interpreted as "the degree to which y belongs to A". The collection of all $\mathbf{L}$-sets in universe Y is denoted by L^Y. $A \in L^Y$ is called crisp if $A(y) \in \{0, 1\}$ for all $y \in Y$. In that case, we may identify a crisp $\mathbf{L}$-set in Y with an ordinary subset of Y. By a slight abuse of notation, if A is crisp, we may write $y \in A$ and $y \notin A$ to denote that $A(y) = 1$ and $A(y) = 0$, respectively. In each universe Y, we consider two borderline (crisp) $\mathbf{L}$-sets: 0_Y such that $0_Y(y) = 0$ for all $y \in Y$ (an empty $\mathbf{L}$-set in Y); 1_Y such that $1_Y(y) = 1$ for all $y \in Y$. If Y is clear from the context, we write $\emptyset$ and Y instead of 0_Y and 1_Y, respectively.

Operations with $\mathbf{L}$-sets we use in this paper are induced componentwise by the operations of $\mathbf{L}$. For instance, the intersection of $\mathbf{L}$-sets $A, B \in L^Y$ is an $\mathbf{L}$-set $A \cap B \in L^Y$ such that $(A \cap B)(y) = A(y) \wedge B(y)$ for all $y \in Y$. The structure of all $\mathbf{L}$-sets in Y together with the induced operations is in fact a direct power $\mathbf{L}^Y = \langle L^Y, \cap, \cup, \otimes, \rightarrow, \emptyset, Y\rangle$ of $\mathbf{L}$ which is also a complete residuated lattice. Moreover, for $A, B \in L^Y$, we define the degree of inclusion of A in B as follows:

$$S(A, B) = \bigwedge\nolimits_{y \in Y} \big(A(y) \rightarrow B(y)\big). \tag{4}$$