Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

The joy of implications, aka pure Horn formulas: Mainly a survey

Marcel Wild

ARTICLE INFO

Article history:
Received 18 March 2015
Received in revised form 4 March 2016
Accepted 16 March 2016
Available online xxxx

Keywords:
Pure Horn functions and their minimization
Boolean logic
Association rule
Lattice theory
Formal Concept Analysis
Closure system
Convex geometry
Prime implicates
Meet-irreducibles
Universal algebra

ABSTRACT

Pure Horn clauses have also been called (among others) functional dependencies, strong association rules, or simply implications. We survey the mathematical theory of implications with an emphasis on the progress made in the last 30 years.

1. Extended introduction

This article is devoted to the mathematics and (to lesser extent) algorithmics of implications; it is mainly a survey of results obtained in the past thirty years but features a few novelties as well. The theory of implications mainly developed, often under mutual ignorance, in these five fields:

Boolean Function Theory, Formal Concept Analysis, Lattice Theory, Relational Database Theory, Learning Theory.

As standard text-books in these fields we recommend [19,33] and [10,31] and [47,44] and [57(ch. VI), 29] respectively. Broadly speaking we collect from each field only those major results that concern (or can be rephrased in terms of) "abstract implications", and *not* the substance matter of the field itself. There are three minor exceptions to this rule. First, there will be two detours (Subsections 4.1, 4.2) into lattice theory; among the five fields mentioned this is the one the author is most acquainted with. Second, in Subsection 1.1 just below, in order to motivate the theory to come, we glance at three "real life" occurrences of implications in these areas: Relational Databases, Formal Concept Analysis, and Learning Spaces. The third exception concerns 3.6; more on that later. The second part (1.2) of our extended introduction gives the detailed section break up of the article.

1.1. We shall only give very rudimentary outlines of three areas mentioned above; more detailed accounts of 1.1.1 to 1.1.3 are found in [47,33,29]. The sole purpose here is to convey a feeling for the many meanings that a statement "A implies B" can have. This will contrast with the uniform mathematical treatment that all "abstract" implications $A \to B$ obey.

E-mail address: mwild@sun.ac.za.

http://dx.doi.org/10.1016/j.tcs.2016.03.018
0304-3975/

Please cite this article in press as: M. Wild, The joy of implications, aka pure Horn formulas: Mainly a survey, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.03.018

	a_1	a_2	a_3	...	...	a_n
t_1	v_{11}	v_{12}	v_{13}	...	...	v_{1n}
t_2	v_{21}	v_{22}	v_{23}	...	...	v_{2n}
⋮	⋮	⋮	⋮			⋮
t_m	v_{m1}	v_{m2}	v_{m3}	...	...	v_{mn}

Fig. 1. Relational database table.

1.1.1. As to relational database (RDB), imagine this as a large array in which every row (called record) corresponds to a particular object t_i, and in which the columns correspond to the various attributes a_j that apply. See Fig. 1. Each attribute has a domain which is the set of values that it may assume. Following an example of J. Ullman, take a relational database whose records match the "teaching events" occurring at a university in a given semester. The attributes are $C =$ course, $T =$ teacher, $H =$ hour, $R =$ room, $S =$ student. The domain of C may be {algebra, analysis, lattice theory, $\cdots$}, the domain of T could be {Breuer, Howell, Janelidze, $\cdots$}, and so forth. If A, B are sets of attributes then the validity of $A \to B$ means that *any two objects which have identical values for all attributes in A, also have identical values for all attributes in B*. Examples of implications $A \to B$ (also called *functional dependencies*) that likely hold in a well designed database include the following: $\{C\} \to \{T\}$ (each course has one teacher), $\{H, R\} \to \{C\}$ (only one course meets in a room at one time), $\{H, S\} \to \{R\}$ (a student can be in only one room at a given time).

1.1.2. Let now G and M be any sets and $I \subseteq G \times M$ be a binary relation. In Formal Concept Analysis (FCA) one calls the triple (G, M, I) a *context*, and gIm is interpreted as the *object* $g \in G$ having the *attribute* $m \in M$. If $A, B \subseteq M$ then the validity of $A \to B$ has a different[1] ring from before: *Any object that has all attributes in A, also has all attributes in B* (see also 2.1.2 and 2.2.3).

Let us focus on particular contexts of type $(G, M, \ni)$. Thus the objects $g \in G$ become *subsets* X of some set M of *items*. Saying that $g \in G$ "has attribute" $m \in M$ now just means $X \ni m$. Often the sets X are called *transactions*, and the elements $m \in M$ are called *items*. If $A, B \subseteq M$ then $A \to B$ is a valid implication iff *every transaction X that contains the itemset A, also contains the itemset B*. For instance, each transaction can contain the items a customer bought at a supermarket on a particular day. In this scenario a plausible implication e.g. is {butter, bread} $\to$ {milk}. Notice that $A \to B$ may be a valid implication simply because many transactions do not contain A at all. To exclude this possibility one often strengthens the previous definition of "valid implication" by additionally demanding that say 70% of all transactions must contain the itemset A. The terminology "transaction" and "itemset" is borrowed from Frequent Set Mining (FSM), a paradigm that developed in parallel to FCA for a long time, despite of close ties. See also 3.6.3.4.

1.1.3. As to Learning Spaces [29], these are mathematical structures applied in mathematical modeling of education. In this framework (closer in spirit to [33] than to [57] type learning theory) the validity of an implication $A \to B$ means the following: *Every student mastering the (types of) problems in set A also masters the problems in set B*. See also Expansion 16.

1.2. Some readers may have guessed that this zoo of implications fits the common hat of pure Horn functions, i.e. Boolean functions like $(x_1 \wedge x_2 \wedge x_3) \to x_4$ and conjunctions thereof. While this is true the author, like others, has opted for a more *stripped down formalism*, using elements and sets rather than literals and truth value assignments, etc. Nevertheless, discarding pure Horn function terminology altogether would be short-sighted; certain aspects can only be treated, in any sensible way, in a framework that provides immediate access to the empire of general Boolean function theory that e.g. houses prime implicates and the consensus algorithm.

Without further mention, all structures considered in this article will be **finite**. Thus we won't point out which concepts extend or can be adapted to the infinite case. A word on [19, chapter 6, 56 pages] is in order. It is a survey on Horn functions to which the present article (PA) compares as follows. Briefly put, the intersection $CH \cap PA$ is sizable (though not notation-wise), and so are $CH \backslash PA$ (e.g. applications, dualization, special classes), as well as $PA \backslash CH$ (e.g. 3.6 and 4.1 to 4.4). We note that 4.1 also features special classes but *others*.

Here comes the section break up. Section 2 recalls the basic connections between closure operators c and closure systems $\mathcal{F}$ (2.1), and then turns to implications "lite" in 2.2. Crucially, each family Σ of implications $A \to B$ gives rise to a closure operator $c(\Sigma, -)$ and whence to a closure system $\mathcal{F} = \mathcal{F}(\Sigma)$. Furthermore, *each* closure operator c is of type $c = c(\Sigma, -)$ for suitable Σ. Section 3 is devoted to the finer theory of implications. Centerpieces are the Duquenne–Guigues implicational base (3.2) and the canonical direct base in 3.3. Subsection 3.4 is about mentioned pure Horn functions, 3.5 is about acyclic and related closure operators, and 3.6 surveys the connections between two devices to grasp closure systems $\mathcal{F}$. One device is any implicational base, the other is the subset $M(\mathcal{F}) \subseteq \mathcal{F}$ of meet-irreducible closed sets.

[1] One may view a context as a RBD all of whose attribute domains are Boolean, thus {True, False} or {1, 0}. But depending on viewing it as RBD or context, different implications hold.

Please cite this article in press as: M. Wild, The joy of implications, aka pure Horn formulas: Mainly a survey, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.03.018