Available online at www.sciencedirect.com
SCIENCE DIRECT®
Discrete Applied Mathematics 147 (2005) 43 – 55
DISCRETE APPLIED MATHEMATICS
www.elsevier.com/locate/dam

Pseudo-models and propositional Horn inference

Bernhard Ganter*, Rüdiger Krauße

Institut für Algebra, Technische Universität Dresden, Zellescher Weg 12-14, D-01062 Dresden, Germany

Received 4 May 2001; received in revised form 6 January 2003; accepted 21 June 2004
Available online 29 December 2004

Abstract

A well-known result is that the inference problem for propositional Horn formulae can be solved in linear time. We show that this remains true even in the presence of arbitrary (static) propositional background knowledge. Our main tool is the notion of a cumulated clause, a slight generalization of the usual clauses in Propositional Logic. We show that each propositional theory has a canonical irredundant base of cumulated clauses, and present an algorithm to compute this base.

MSC: 03B05; 03B35; 68T30

Keywords: Horn inference; Horn base; Background knowledge

1. Introduction

Inferring information can be a tedious task, even in the simple setting of Propositional Logic: to decide if a propositional formula follows from a given list of formulae is, in general, an $\mathcal{NP}$-complete problem. If, however, all the formulae under consideration are implications, then the inference problem can easily be solved.

Implications are particularly natural to describe classifications, when objects are grouped with regard to selected attributes. In such a situation, implications encode expressions of the form "each object with attributes $a_1, \ldots, a_n$ also has attributes $b_1, \ldots, b_m$". The simple

* Corresponding author.
E-mail addresses: ganter@math.tu-dresden.de (B. Ganter), mail@krausze.de (R. Krauße).

0166-218X/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2004.06.019

inference mechanism allows to effectively study the system of all implications that hold in a given situation and to construct an irredundant implicational base, see [4, Section 2.3].

Implications are natural to describe classifications because objects are usually grouped according to their common attributes, whereas disjunctions and negations are of lesser importance. Nevertheless, for some purposes implications are insufficient: they cannot express that certain attributes exclude each other or that the absence of an attribute implies the presence of another. In a knowledge acquisition process, such nonimplicational information is often static and has the character of background information. For many-valued attributes (see [4, Section 1.3]), such background knowledge may describe the implicit structure of the attribute values.

An implicational knowledge acquisition method with propositional background knowledge has been described in [3], with emphasis on the methodology rather than on the efficiency. The algorithmic problems have been studied in [7]. The present paper combines and extends results from these. To make it accessible to a wider audience it is formulated in the language of Propositional Logic, in contrast to its predecessors, which use the terminology of Formal Concept Analysis.

We proceed as follows: first we introduce the notion of a *pseudo-model*. This can be done without reference to logic, in terms of elementary set theory. Next, we define a class of propositional formulae which we call *cumulated clauses* (because they are conjunctions of clauses having the same negated part). Cumulated clauses generalize the usual clauses; therefore each propositional theory is generated by its cumulated clauses. We prove an "irredundant basis" result for cumulated clauses and give an algorithm to construct such a base. Unfortunately, unlike in the case of implications, this basis is not necessary of minimal cardinality. Our final result shows that the complexity of implicational inference, modulo background knowledge given in the form of cumulated clauses, depends linearly on the implicational part.

It is due to space limitations that we illustrate our results below only by toy examples. Serious applications exist and will be published elsewhere. The reader is referred to [3] for a first impression.

2. Pseudo-models

Let M be a finite set and let $\mathscr{F}$ be a set of some subsets of M. For reasons that will become transparent later we call the elements of $\mathscr{F}$ the "models".

Definition 1. A set $P \subseteq M$ is a *pseudo-model* of $\mathscr{F} \subseteq \mathfrak{P}(M)$ if

(1) $P \notin \mathscr{F}$, and
(2) for each pseudo-model $Q \subseteq P$, $Q \neq P$, there is some $F \in \mathscr{F}$ with $Q \subseteq F \subseteq P$.

This is of course a recursive definition: the set $\emptyset$ is a pseudo-model if and only if $\emptyset \notin \mathscr{F}$, and as soon as the pseudo-models of cardinality less than n are known the definition specifies which sets of cardinality n are pseudo-models.