Journal of Computer and System Sciences 81 (2015) 208–218

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Decision implication canonical basis: a logical perspective

Yanhui Zhai, Deyu Li *, Kaishe Qu

School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

ARTICLE INFO

Article history:
Received 11 October 2011
Received in revised form 19 March 2014
Accepted 16 May 2014
Available online 11 June 2014

Keywords:
Formal concept analysis
Decision implication
Decision context
Canonical basis

ABSTRACT

Due to its special role on logical deduction and practical applications of attribute implications, canonical basis has attracted much attention and been widely studied in Formal Concept Analysis. Canonical basis is constructed on pseudo-intents and, as an attribute implication basis, possesses of many important features, such as completeness, non-redundancy and minimality among all complete sets of attribute implications. In this paper, to deduce an analogous basis for decision implications, we introduce the notion of decision premise and form the so-called *decision implication canonical basis*. Furthermore, we show that the basis is complete, non-redundant and minimal among all complete sets of decision implications. We also present an algorithm to generate this canonical basis and analyze time complexity of this algorithm.

1. Introduction

R. Wille [26] introduced Formal Concept Analysis (FCA) as an order-theoretic method for the mathematical analysis of binary data. The starting point of FCA is based on a formalization of the philosophical understanding of a *concept* as a unit of thought constituted by its *extent* and *intent*. The extent of a concept is understood as the set of all objects belonging to the concept and the intent as the multitude of all attributes common to all those objects. The transformation from two-dimensional incidence tables to *concept lattices* structure is a crucial keystone from which FCA derives much of its power and versatility as a modeling tool. The concept lattices obtained the way turn out to be exactly the complete lattices, and the particular way in which they structure and represent knowledge is very appealing and natural from the perspective of many scientific disciplines. Over the past thirty years, FCA has been widely studied [9,8,20,24,4] and become a powerful tool for machine learning [11,29], software engineering [24] and information retrieval [4].

One of the aspects of FCA is attribute logic on attribute implication [22,9]. In FCA, an attribute implication is of the form $A \to B$, meaning that one can derive B from A. Then an attribute implication is *valid* in a set of data, if no data violates the attribute implication. A set of valid attribute implications may also be *complete* and *non-redundant*. A complete set of attribute implications carries all information from the data and thus one can reconstruct the data from the set, whereas a non-redundant set is a compact representation of attribute implications, indicating that the attribute implications in the set are independent of one another.

Concerning complete sets of attribute implications, [19] provided a complete, but redundant set of attribute implications, and described an algorithm for generating the set with a NextClosure-based algorithm [7]. More well-known is a complete and non-redundant set generated from *proper premises* [9], which produces its premises by removing some redundant information from closures. Pseudo-intent, introduced by V. Duquenne and J.L. Guigues [6], has obtained wide interests [9,12,2],

* Corresponding author.
E-mail addresses: chai_yanhui@163.com (Y. Zhai), lidysxu@163.com (D. Li), quks@sxu.edu.cn (K. Qu).

http://dx.doi.org/10.1016/j.jcss.2014.06.001

Table 1
A formal context.

	a_1	a_2	a_3	a_4	a_5	a_6	d_1	d_2
x_1	×	×			×	×	×	
x_2			×					×
x_3	×		×			×	×	
x_4		×						
x_5		×	×	×	×	×	×	×
x_6			×		×			
x_7		×	×	×		×		×
x_8	×	×		×	×	×	×	

because it corresponds to an optimal representation of implications. Canonical basis, whose premises are exactly pseudo-intents, is proven to be complete, non-redundant, and more importantly, minimal among all complete sets of implications. However, finding a pseudo-intent is not an easy problem. There are some open problems [18] concerning the complexity of generating and finding pseudo-intents. For example, it has been proven that checking whether a subset is a pseudo-intent is coNP-complete [15,2], and that counting the number of pseudo-intents is even #P-hard [12,14]. To overcome this problem, Obiedkov etc. [16] showed some "genealogic" properties of attribute implications and presented an attribute-incremental algorithm for computing canonical basis. Experiment results showed that this algorithm is quite competitive. Valtchev etc. [25] adapted the divide-and-conquer policy and presented a method for computing canonical basis, which outperformed NextClosure on some datasets.

On the other hand, decision-based FCA (including decision context and decision implication) has been widely studied [10,11,13,20,28,27]. In the literature [20], Qu etc. presented a special inference rule, called α-decision inference rule, which may deduce other decision implications by enlarging premises of implications and/or reducing corresponding consequences. In the setting, [20] obtained an α-complete and α-non-redundant set of decision implications and showed that the complete set can be characterized by minimal generators [21,23,5]. In addition, an algorithm for generating the complete set was then given based on minimal generators and NextClosure algorithm. Afterwards, Zhai etc. [27] formulated decision implications and presented logical characteristics of decision implications. Specifically, [27] introduced the notions of *closure* and *unite closure* and established the semantical aspect of decision implications; [27] then formed two deduction rules and showed that the two rules are complete with respect to the semantical aspect, which established the syntactical aspect of decision implications.

Following [20] and [27], the paper intends to construct a canonical basis for decision implications. This canonical basis, called decision implication canonical basis is with the so-called *decision premises* (*d-premises*) as its premises of decision implications and closures on decision subcontext as its consequences of decision implications. Decision implication canonical basis is semantically complete and non-redundant, and furthermore it contains the least number of decision implications among all complete subsets of implications. In other words, d-premise is a counterpart of pseudo-intent in the case of decision implications.

This paper is organized as follows. Section 2 presents some basic notions about FCA and decision contexts, which are taken from [9,13,20]. We reformulate decision implication in terms of logic in Section 3. In Section 4, we introduce the notion of decision premise and prove that the so-called decision implication canonical basis is complete and non-redundant. Besides, we also show that the canonical basis contains the least number of decision implications among all complete subsets of implications. Section 5 discusses how to generate d-premises and decision implication canonical basis. Section 6 concludes the paper and lists some further remarks.

2. Formal concept analysis

2.1. Basic notions of FCA

This subsection provides a brief overview of FCA, and for more extensive introduction refer to [9].

A triple $K = (G, M, I)$ is a formal context, if G and M are sets, and $I \subseteq G \times M$ is a binary relation. In the case, the elements of G are called objects, the elements of M are called attributes, and I is viewed as an incidence relation between objects and attributes.

Example 1. Formal contexts are mostly represented by rectangular tables and an example is illustrated by Table 1. In the table, a cross means that the row object has the column attribute.

Within formal context, we can define some operators on object subsets and attribute subsets. Specifically, for a subset $A \subseteq G$ of objects we define:

$$A^I = \{m \in M \mid gIm, \forall g \in A\}$$