ARTICLE IN PRESS

Journal of Mathematical Psychology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

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

Compressed representation of Learning Spaces

Marcel Wild

Department of Mathematical Sciences, University of Stellenbosch, Matieland, 7600, South Africa

HIGHLIGHTS

- The size of an implication base of a Learning Space can be bounded by the square of the number of join-irreducible knowledge states.
- New compression methods based on wildcards are proposed both for Knowledge Spaces and for Learning Spaces.
- We compare, and enhance, the query-learning techniques used in Formal Concept Analysis and Learning Space Theory respectively.

ARTICLE INFO

Article history:
Received 24 April 2015
Received in revised form
24 February 2017
Available online xxxx

Keywords:
Learning Space
Data compression
Pure Horn formula
Query learning
Formal Concept Analysis
Antimatroid

ABSTRACT

Learning Spaces are certain set systems that are applied in the mathematical modeling of education. We propose a wildcard-based compression (without loss of information) of such set systems to facilitate their logical and statistical analysis. Under certain circumstances compression is the prerequisite to calculate the Learning Space in the first place. There are connections to the dual framework of Formal Concept Analysis and in particular to so called attribute exploration.

1. Introduction

In order to grasp the structure of this article one needs a basic understanding of what Knowledge Spaces and the more specific Learning Spaces are all about. Our introductory example in 1.1, that aims to convey the gist of Learning Spaces to the novice, is based on the introductory example of Doignon (2014). Only afterwards we will be in a position to state our main contribution (in 1.2), and to proceed with the Section break up (in 1.3).

1.1

A 'knowledge structure' $(Q, \mathcal{K})$ consists of a domain Q and a collection $\mathcal{K}$ of subsets K of Q. The elements of Q, also called *items*, are the elementary pieces of information. Each subset K, also called *knowledge state*, contains all the items mastered by some hypothetical student at some given time. In particular, complete ignorance and omniscience are represented respectively, by the empty set and the domain itself. Besides these two extreme knowledge states, $\mathcal{K}$ usually contains many more (knowledge) states. They correspond to potential ("medium-talented") students. An example $(Q_1, \mathcal{K}_1)$ with domain $Q_1 = \{a, b, c, d\}$ is shown in Fig. 1; the circles display the 9 states, and the lines between them indicate covering states. Let us explain why a knowledge structure like $(Q_1, \mathcal{K}_1)$ is actually unlikely to occur in practice. The state $\{c, d\}$ belongs to $\mathcal{K}_1$ but it is impossible for a learner to acquire the mastery of c and d one after the other. This is because neither $\{c\}$ nor $\{d\}$ is knowledge state. This goes against the common view that learning happens step-wise, one item a time. As another counter-intuitive phenomenon, consider a student in state $\{b\}$. He many learn item c to reach state $\{b, c\}$. Likewise, while in state $\{b\}$, he may decide to first learn a, which brings him to state $\{a, b\}$. But then, strangely enough, item c is no longer learnable since $\{a, b, c\}$ is no knowledge state.

The definition of a 'Learning Space' as a particular type of knowledge structure avoids the two strange scenarios that we just illustrated. It imposes the following two conditions on the states of a knowledge structure $(Q, \mathcal{K})$.

(A) ACCESSIBILITY. Any state K contains an item q such that $K \setminus \{q\}$ is again a state.

(LC) LEARNING CONSISTENCY. For a state K and items q, r if $K \cup \{q\}$ and $K \cup \{r\}$ are states, then $K \cup \{q, r\}$ is also a state.

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

http://dx.doi.org/10.1016/j.jmp.2017.02.002
0022-2496/

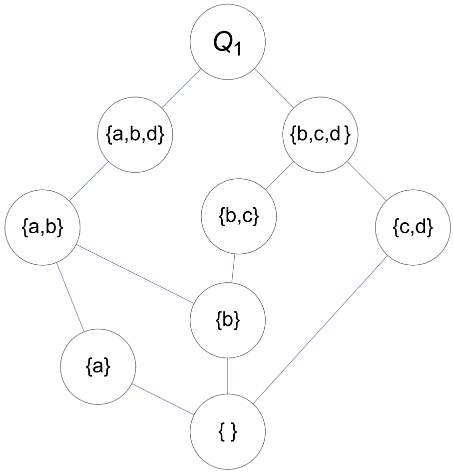

Fig. 1. A "bad" knowledge structure $\mathcal{K}_1$.

1.2

The primary purpose of this article is the application of compression techniques (previously explored by the author in other contexts) to accommodate knowledge structures with millions of states. That not only reduces storage space but also facilitates statistical analysis. In the framework of the more specific Learning Spaces these compression techniques naturally lead to "query learning" which constitutes the second theme of our article. Although the author's expertise is skewed towards the first theme, the research directions proposed for the second are deemed to be fruitful.

1.3

Here comes the section break up. Section 2 introduces, by way of a toy example, the basic idea of how large chunks of the powerset $\mathcal{P}(Q)$ can be chopped away in such a way that the desired knowledge structure $\mathcal{K}$ results in a compressed fashion.

Section 3 presents both well and lesser known facts about specific knowledge structures, i.e. so called Knowledge Spaces $(Q, \mathcal{K})$. In 3.1 we introduce the base $\mathcal{B}(\mathcal{K}) \subseteq \mathcal{K}$. This leads (3.2) to Dowling's algorithm that generates $\mathcal{K}$ from $\mathcal{B}(\mathcal{K})$. In 3.3 we make precise the informal "dual implications" (= dimplications) occurring in Section 2. (The matching term in Falmagne and Doignon (2011) is "entailment".) Of particular importance are *prime* dimplications. In Theorem 1 we show how the set PrimeDimp$(\mathcal{K})$ of all prime dimplications of a Knowledge Space $\mathcal{K}$ can be calculated from $\mathcal{B}(\mathcal{K})$. In 3.4 we see that PrimeDimp$(\mathcal{K})$ is just one example (though an important one) of a "dimplication base" of $\mathcal{K}$. Section 3.5 is about Learning Spaces $\mathcal{K}$, as defined by (A) and (LC) above. Learning Spaces are Knowledge Spaces $\mathcal{K}$ for which both $\mathcal{B}(\mathcal{K})$ and PrimeDimp$(\mathcal{K})$ are particularly well-behaved. Section 3.6 elaborates on the well-known fact (Falmagne & Doignon, 2011) that Learning Spaces are known as antimatroids in the Combinatorics and Operations Research communities.

Section 4 is in the spirit of Section 2 but with more sophisticated don't-care symbols (aka wildcards). The underlying *e-algorithm* was previously applied by the author in other circumstances. Here we show that, given any base Θ of dimplications of an (unknown) Knowledge Space $\mathcal{K}$, the *e*-algorithm can calculate a compact representation of $\mathcal{K}$. In 4.2 the latter is used for statistical analysis (as alluded to in 1.2), and in 4.3 we show how the base $\mathcal{B}(\mathcal{K})$ can be sieved from it.

Section 5 recalls the duality between Knowledge Spaces $\mathcal{K}$ and closure spaces $\mathcal{C}$. In particular, dimplications correspond to the better known implications. In 5.1 we introduce lattices and show how each lattice $\mathcal{L}$ can be modeled naturally by a closure system $\mathcal{C}(\mathcal{L})$. This allows to apply the theory of implications to lattices. Many specific lattices have been investigated in this regard, see Wild (2017) for a survey. For our purpose so called meet-semidistributive lattices come into focus; the relevant facts are readied in 5.2.

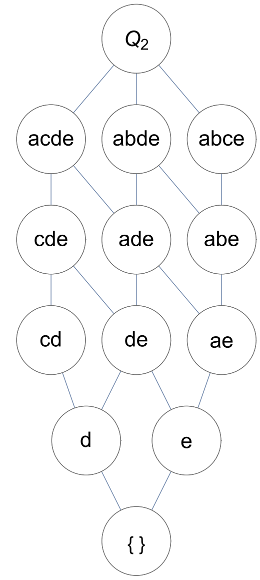

Fig. 2. A "good" knowledge structure $\mathcal{K}_2$.

This is exploited in Section 6 where it leads to a second method to compress a Learning Space, apart from the way in Section 4 which works for any Knowledge Space. In brief, whereas the *e*-algorithm from Section 4 operates on the universe Q, the *n*-algorithm from Section 6 has $\mathcal{B}$ as its universe, and always $|\mathcal{B}| \geq |Q|$. As opposed to Θ in Section 4, the size of the base Σ of implications derived from $\mathcal{B}$ can be bounded as $|\Sigma| \leq |\mathcal{B}|^2$ by Theorem 2.

Section 7 evaluates the discussed algorithms on computer-generated random examples.

Section 8 dwells on the "query learning" aspect of it all. Thus the framework of Formal Concept Analysis will be compared to Knowledge Space Theory, and we glimpse at the general theory of "learning Boolean functions".

2. Compression of knowledge structures using don't-care symbols

A dual implication or briefly "dimplication" is a certain statement about a knowledge structure which is either true or false. To fix ideas, consider the knowledge structure $\mathcal{K}_2$ in Fig. 2 (which is based on Fig. 15.1 in Falmagne & Doignon, 2011) with domain $Q_2 = \{a, b, c, d, e\}$. By definition the dimplication $\{b, d\} \rightsquigarrow c$ "holds" in $\mathcal{K}_2$ when every student who fails both b and d also fails c. Put another way: The mastering of c implies the mastering of b or d. If $\mathcal{K}_2$ is *known* in one way or another, e.g. in diagram form as in Fig. 2, then it is easy in principle (but possibly tiresome in practice) to decide whether some dimplication holds. In our case $\{b, d\} \rightsquigarrow c$ holds in $\mathcal{K}_2$ because (check) every knowledge state $K \in \mathcal{K}_2$ that contains c also contains b or d. Likewise $\{b, c\} \rightsquigarrow e$ does *not* hold in $\mathcal{K}_2$ because (say) $K = \{a, e\}$ contains e but neither b nor c.

2.1

Before continuing with our toy example $\mathcal{K}_2$ it pays to properly formalize dimplications. This concept was introduced in Koppen