

Combinatorics in Education: An Introduction to Knowledge Space Theory

Carson Mulvey

MAD4204

Introduction

The goal of this project is to give an introductory overview to the theory of knowledge and learning spaces, a field that mixes combinatorics and mathematical psychology. We will cover some principles of knowledge and learning spaces and demonstrate algorithmic solutions to creating these spaces. We will also show the connections between knowledge space theory and the content from our course. Learning spaces in particular closely tie into poset and lattice theory, and as such, the knowledge from course content relating to posets is assumed. The active component of the project is an implementation one of the explained algorithms. The source code is available on GitHub ¹ and a demonstration is available on YouTube ².

The theory of knowledge spaces is relatively new, beginning with a joint paper between Doignon and Falmagne (1985). The field uses a combinatorial approach to assessing competency in a skill, attempting to create more informative psychometric measures than the singular numbers produced by assessments such as the SAT and ACT. Since then, the field has been extensively researched and applied by the popular website **A**ssessment and **L**earning in **K**nowledge **S**paces, better known as ALEKS, to develop an

¹<https://gist.github.com/heyuncle/94b858fdbb9a256d424766b361e6c283>

²<https://youtu.be/gebNB3qwiBQ>

online platform for learning math and chemistry. The theory of knowledge spaces is still currently researched and applied to online platforms like ALEKS to more effectively teach and measure skill levels in various school subjects.

1 Knowledge Space Theory

Suppose that as a teacher, we want to measure a student's ability in a subject. To have a detailed understanding of what the student does or doesn't understand, we split a subject into a set of distinct 'skills' Q . Then for some $K \subseteq Q$, we can think of K as the skills that a student already knows. The empty set and Q itself are clearly possibilities for K , representing the student knowing nothing or everything about the subject, respectively. However, not all K are necessarily *feasible* subsets of skills a student can have, leading to our first definition.

Definition 1.1. The pair (Q, \mathcal{K}) is a *knowledge structure*, given nonempty set Q and family $\mathcal{K} \subseteq 2^Q$ containing Q and \emptyset . We call Q the *domain* and \mathcal{K} the (*knowledge*) *states* of the knowledge space.

Note that because $\bigcup_{K \in \mathcal{K}} K = Q$, we may denote a knowledge structure by just \mathcal{K} without causing any ambiguity.

Also, we often use \mathcal{K}_q for some $q \in Q$ to denote the subset of \mathcal{K}_q of all knowledge states containing q .

Example 1.1. Consider the domain $Q = \{a, b, c, d\}$ with knowledge states

$$\mathcal{K} = \{\emptyset, \{a\}, \{b\}, \{a, b\}, \{a, c\}, \{a, d\}, \{a, b, c\}, \{a, b, d\}, \{a, c, d\}, Q\}.$$

We see that (Q, \mathcal{K}) is indeed a knowledge structure, as all subsets $K \in \mathcal{K}$ are in the power set of Q , with both \emptyset and Q itself in \mathcal{K} .

However, this definition alone is very loose; an arbitrary subset of skills K on its own encodes very little pedagogical information. As such, we add restrictions that lead to the *knowledge spaces* and *learning spaces* that make up the majority of the theory.

Definition 1.2. A knowledge structure (Q, \mathcal{K}) (also commonly notated as \mathcal{K}) is called a *knowledge space* if it is *union-closed*, i.e. if for all $A, B \in \mathcal{F} \subseteq \mathcal{K}$, we have $A \cup B \in \mathcal{F}$.

A *quasi-ordinal knowledge space* is a knowledge space \mathcal{K} with the additional restriction of being *intersection-closed*, i.e. if for all $A, B \in \mathcal{F} \subseteq \mathcal{K}$, we have $A \cap B \in \mathcal{F}$.

Knowledge spaces, as opposed to knowledge structures, add a meaningful restriction in terms of pedagogical interpretation. They can be interpreted as the existence of an expert on a topic in a class allowing for all students to eventually master that topic as well. Additionally, a quasi-ordinal knowledge structure has the additional property that if one student knows skills a and b , and another student knows skills b and c , then a third student may only know skill b .

Definition 1.3. A *well-graded knowledge space* is a knowledge space \mathcal{K} satisfying the additional axiom:

(MA) For nonempty $K \in \mathcal{K}$, there exists some q in K such that $K \setminus \{q\} \in \mathcal{K}$.

Well-graded knowledge spaces allow us to create a ‘tight path’ between two knowledge states, i.e. a sequence $K = K_0, K_1, \dots, K_p = L$ where $|A \triangle B| = 1$ for $i \in [p]$. Also, a union-closed family satisfying the (MA) axiom may also be referred to as an *antimatroid*.

Finally, we define a very special case of a knowledge space called a learning space.

Definition 1.4. A knowledge structure (Q, \mathcal{K}) is a *learning space* if it satisfies the following axioms:

(L1) For any two states K and L with $K \subset L$, there exists a finite chain

$$K = K_0 \subset K_1 \subset \dots \subset K_p = L$$

with $|K_i \setminus K_{i-1}| = 1$ for $i \in [p]$.

(L2) For any two states K and L with $K \subset L$, if some q satisfies $K + \{q\} \in \mathcal{K}$, then $L \cup \{q\} \in \mathcal{K}$.

Axiom (L1) can be interpreted as the ability to go between knowledge states by learning skills one at a time. Axiom (L2), on the other hand, shows that a learner can continue learning even at a greater knowledge state.

Finally, the following result connects these varieties of knowledge structures together.

Theorem 1.1. For a knowledge structure (Q, K) , the following statements are equivalent:

1. (Q, K) is a learning space.
2. (Q, K) is a well-graded knowledge space.

2 Building Knowledge Spaces: Naive QUERY Algorithm

In the context of knowledge spaces, we often want to create a knowledge space by finding the feasible knowledge states for a particular subject. However, to make these states accurate, we must find a heuristic that will let us determine when a state is **not** feasible. As such, we create the following query:

- (Q1) Suppose that a student has failed to solve all items in a nonempty subset A of domain Q . Do you believe this student would also fail to solve item q ? You may assume that chance factors, such as lucky guesses and careless errors, do not interfere in the student's performance.

This query was developed for the **QUERY** algorithm, an algorithm developed by Koppen in 1993. We will show the process of a simplified version of **QUERY**, known by Doignon and Falmagne as 'a naive querying algorithm'.

First, we will define a relevant relation, an 'entailment' for a domain Q .

Definition 2.1. An *entailment* for a domain Q is a relation \mathcal{P} from $2^Q \setminus \{\emptyset\}$ to Q satisfying

$$A\mathcal{P}q \iff (\forall K \in \mathcal{K} : A \cap K = \emptyset \Rightarrow q \notin K).$$

An entailment on Q satisfies the following properties:

- (i) If $q \in A \subseteq Q$, then $A\mathcal{P}q$.
- (ii) If $A, B \in 2^Q \setminus \{\emptyset\}$ and $q \in Q$, then $A\mathcal{P}b$ for all $b \in B$ and $B\mathcal{P}p$ imply $A\mathcal{P}p$.

Most importantly, a theorem from Doignon and Falmagne is used to connect membership in an entailment and feasible states in a knowledge space.

Theorem 2.1. There is a one-to-one correspondence between the family of all knowledge spaces K on the same domain Q , and the family of all entailments \mathcal{P} for Q . This correspondence is defined by the two equivalences

$$A\mathcal{P}q \iff (\forall K \in \mathcal{K} : A \cap K = \emptyset \Rightarrow q \notin K), \quad (1)$$

$$K \in \mathcal{K} \iff (\forall (A, p) \in \mathcal{P} : A \cap K = \emptyset \Rightarrow p \notin K). \quad (2)$$

Using (2) from this equivalence lets us design the following algorithm:

- Step 1. Draw up the list of all the subsets of Q .
- Step 2. Successively submit all the queries (A, q) of the form [Q1]. Whenever a positive response $A\mathcal{P}q$ is observed, remove from the list of remaining subsets all the sets containing q and disjoint from A .

If all queries are submitted, then (2) shows exactly which subsets K are or are not in \mathcal{K} , hence showing the exact knowledge space \mathcal{K} .