Knowledge Spaces: The Mathematics Behind Assessment and Learning

Qingci An, Madina Bolat, José Sanchez, Michael Santana (Adviser), A.J. Hildebrand (Adviser)

ILLINOIS UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN NAZARBAYEV UNIVERSITY

Illinois Summer Research Symposium, July 21, 2016

Background

The theory of knowledge spaces was first introduced by Jean-Claude Falmagne in 1985 in order to build an efficient machine for assessing knowledge.

- This theory was fully developed by Jean-Claude Falmagne and Jean-Paul Doignon.
- Practical application: ALEKS (Assessment and Learning in Knowledge Spaces)
- In 2014, ALEKS had more than 1,900,000 single users.

Knowledge Structures

A *knowledge structure* is a pair $(Q, \mathcal{K})$ with $Q \neq \emptyset$ and $\mathcal{K}$ a family of subsets of Q such that $Q \in \mathcal{K}$ and $\emptyset \in \mathcal{K}$.

- Q is called the *domain* of the knowledge structure;
- Elements of Q are *items* or *concepts*;
- Elements of $\mathcal{K}$ are called *knowledge states*.

Example: Given a test, Q is the set of all concepts that are tested, and a knowledge state is a set of all concepts mastered by a student.

Learning Spaces

A knowledge structure $(Q, \mathcal{K})$ is called a *learning space* if it satisfies the following conditions:

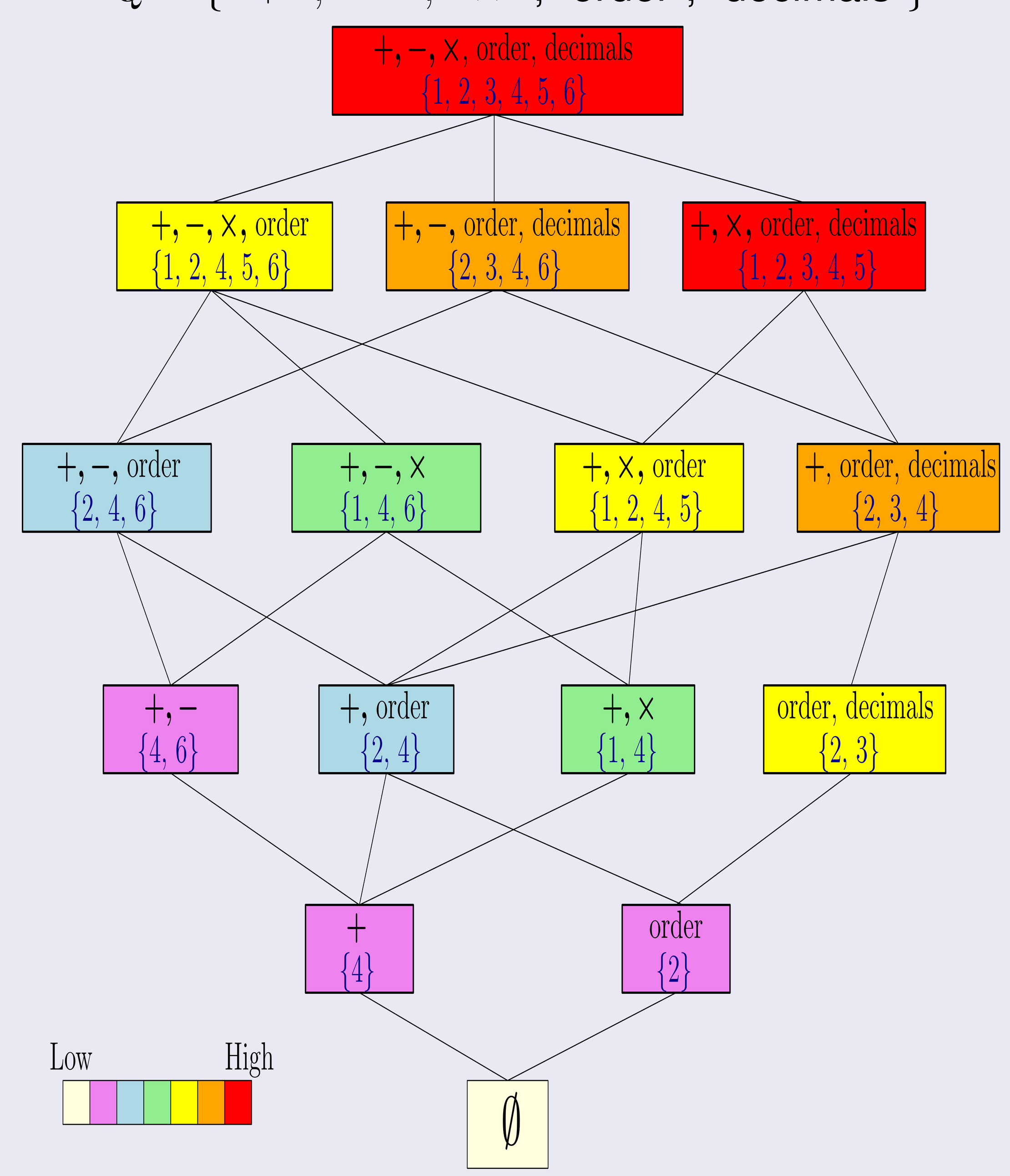
1. **Learning smoothness:** For any states K and L in $\mathcal{K}$ with $K \subset L$, there exists a finite chain of states $K = K_0 \subseteq K_1 \subseteq K1 \subseteq \cdots \subseteq K_p = L$ such that $|K_i \backslash K_{i-1}| = 1$ for $1 \leq i \leq p$ where $p = |L \backslash K|$. (Learning is done one step at a time)
2. **Learning consistency:** For any states K and L in $\mathcal{K}$ with $K \subseteq L$, if q is an item such that $K \cup \{q\}$ is a state, then $L \cup \{q\}$ is also a state. (Knowing more does not prevent learning something new).

NAEP

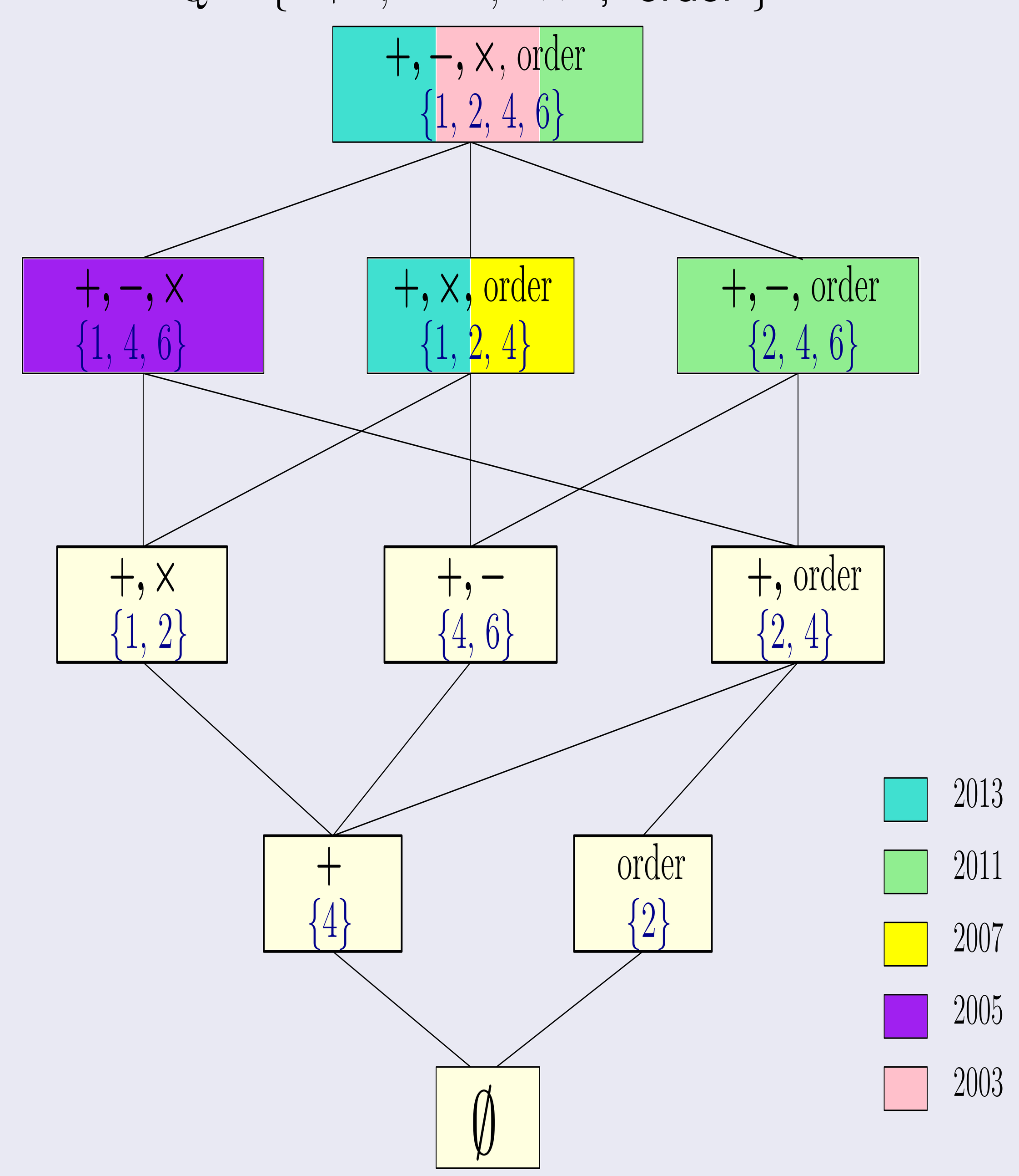
- National Assessment of Educational Progress (NAEP) monitors what American students know in different subject areas like mathematics, history, economics, etc.
- NAEP releases a sample of questions and students' performance data for various years.
- For mathematics the database contains information from assessments of 4th, 8th, and 12th grade students. We considered a set of mathematics questions given to 4^{th} grade students.

Learning Space From Real-World Data

$Q = \{``+", ``-", ``\times", \text{"order"}, \text{"decimals"}\}$



$Q = \{``+", ``-", ``\times", \text{"order"}\}$

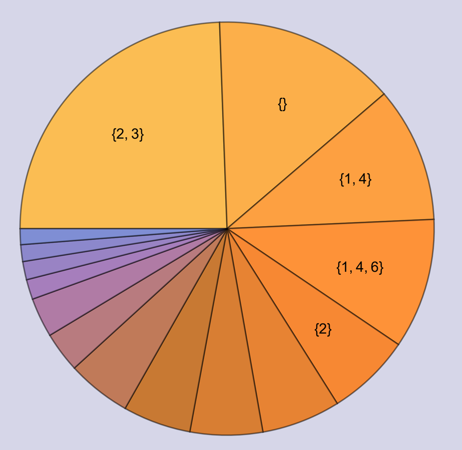


The diagram on the left represents a learning space developed from 2013 data. We found questions corresponding to various items, and using the performance data, we simulated multiple tests to determine which state is dominant. The diagram on the right represents a learning space that spans several years of data and compares their dominant states.

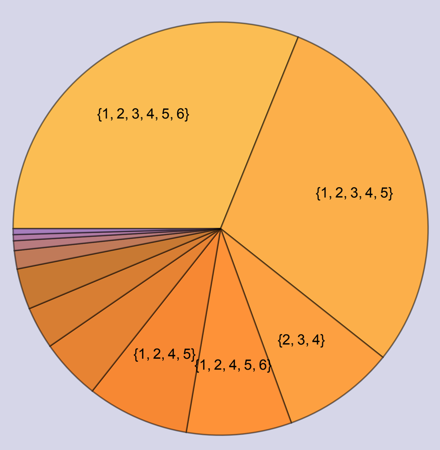
Simulations

Progressive Test

We simulate two tests that progress through questions based on student responses, one with increasing difficulty and the other with decreasing difficulty. These special orderings eliminate non-feasible states.



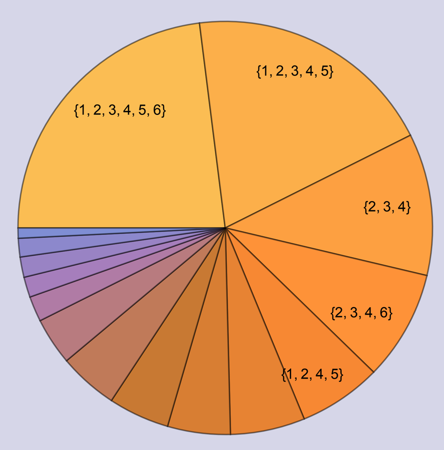
Increasing difficulty



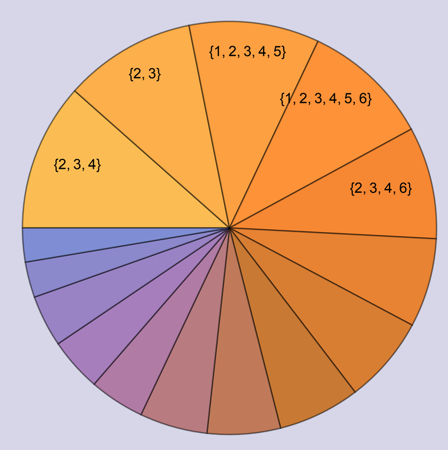
Decreasing difficulty

Probability Variations

Uses a normal distribution to calculate a variation in probabilities of correct responses by students at a certain percentile. Identifies most frequent knowledge state for each percentile and finds when a dominant state changes.



Top 37% students



Top 70% students

Conclusions

1. The knowledge states that differed only by Question 6 had very similar frequencies. We deduced that this occurred because the probability of a correct response to Question 6 is around 50%.
2. There appears to be no trend between years. The dominant state(s) for each year differs from the entire domain by at most one concept.
3. The inconsistency between the most frequent knowledge states of the progressive test in increasing difficulty and the progressive test in decreasing difficulty led us to conclude that the ordering of the questions plays an important role in assessing knowledge.
4. We can predict the performance of a class or a student at a certain percentile using the probability variations simulation.

Future Work

- Write a code that builds a learning space given a domain and precedence relations' information.
- Build a more accurate and efficient simulation that checks a student's mastery of a concept through several questions.

Acknowledgements

This poster was produced as part of the Illinois Summer Mathematics REU program, with support from the Department of Mathematics at the University of Illinois at Urbana-Champaign and Nazarbayev University, Astana, Kazakhstan. The figures were generated by Wolfram Mathematica 10.3 and Ipe Extensible Drawing Editor. The NAEP data was obtained from http://nces.ed.gov/.

References

- J.-P. Doignon, *Learning Spaces, and How to Build Them*, Formal Concept Analysis 2014
- J.-P. Doignon, J.-C. Falmagne, *Knowledge Spaces*, Springer-Verlag 1999
- J.-P. Doignon, J.-C. Falmagne, *Knowledge Spaces and Learning Spaces*, New Handbook of Mathematical Psychology 2015
- J.-P. Doignon, J.-C. Falmagne, *Learning Spaces*, Springer-Verlag 2011