A Data-Driven Framework of Modeling Skill Combinations for Deeper Knowledge Tracing

Yun Huang[1], Julio Guerra[2,3], Peter Brusilovsky[1,2]

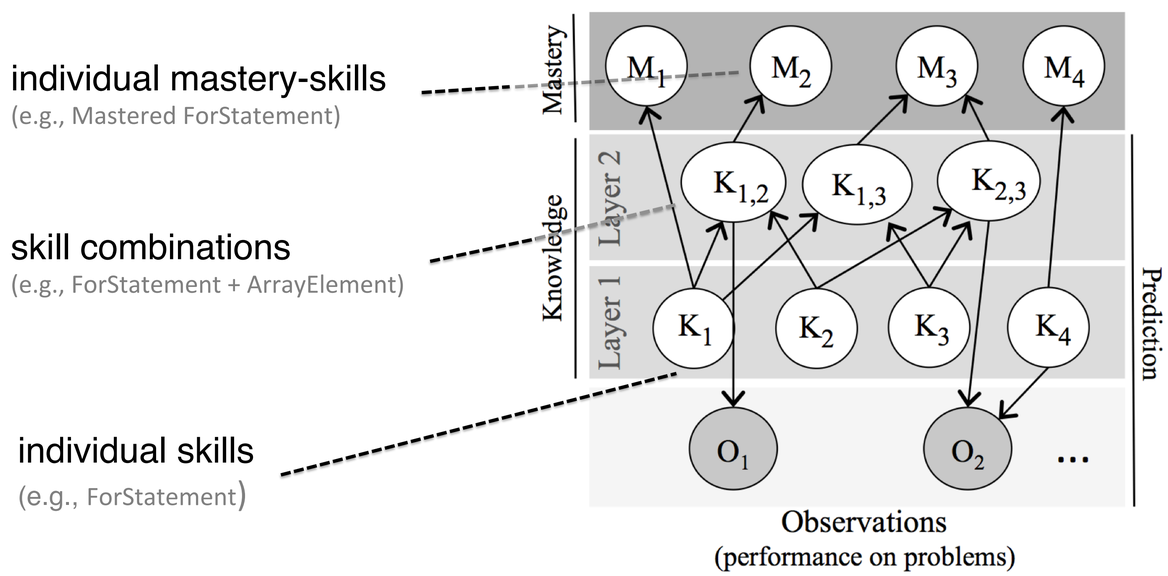
[1] University of Pittsburgh - Intelligent Systems Program , [2] University of Pittsburgh - School of Information Sciences,
[3] Universidad Austral de Chile - Instituto de Informática

Motivation

Combination of skills *matters* for modeling learners: ***mastery* should be asserted only when a student can fluently apply skills in combination with other skills**. However, simply creating such combination skills as new independent skills may ignore the important dependencies for reliable inferences.

Proposed Framework

A data-driven **framework** to model skill combinations called *Conjunctive Knowledge Modeling with Hierarchical Skill Combinations* (CKM-HSC)



Skill combinations (Layer 2) are selected if:

- Skill combination is *much more difficult* than each of its individual skills.
- Skill combination difficulty is *high*.
- *Difficult* problems (items) likely require skill combinations.
- Each problem has a limited number of skill combinations.

Network structure is learned using a greedy search algorithm (we proposed a simplified version using empirical pruning with higher efficiency).

Evaluation

Multifaceted data-driven evaluation framework that includes:

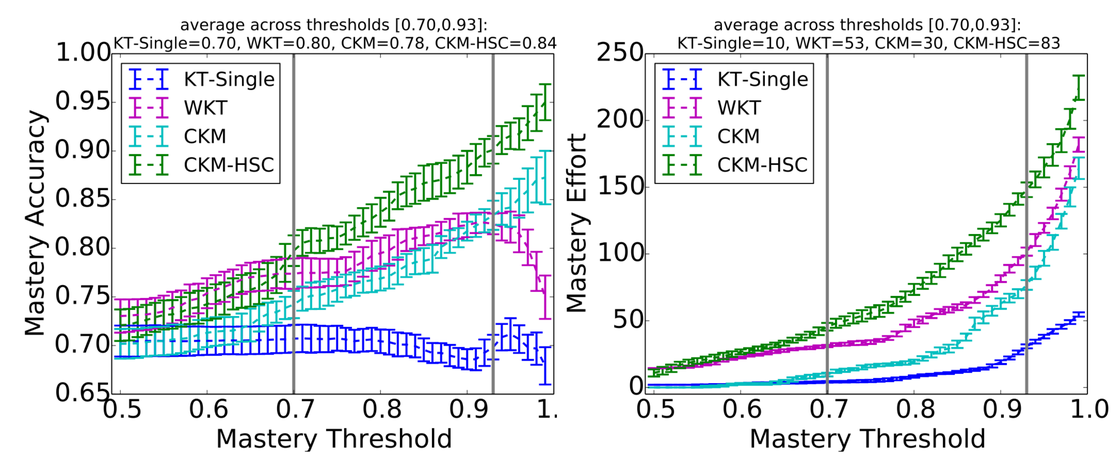
- **Knowledge inference quality:**
  - *Mastery Accuracy*: Do students mostly have correct responses on the data after a student model infers mastery?
  - *Mastery Effort*: How many practices does a student need to reach inferred mastery for all required skills on the data?
- **Parameter plausibility:** Item Discriminative Index (IDI=1-guess-slip)
- **Performance prediction accuracy:** RMSE, AUC

These metrics extend our recent learner effort-outcome paradigm (LEOPARD) [5] and the Polygon multifaceted evaluation framework [7].

Studies

Table 1: Dataset descriptive statistics (from two programming courses at University of Pittsburgh).

Dataset	#obs.	#items	#skills	avg #skills/item	#users	%correct
SQL	17,197	45	34	5 (from 1 to 10)	366	58%
Java	25,988	45	56	5 (from 1 to 11)	347	67%



CKM-HSC significantly **increases the mastery inference accuracy**, and **more reasonably distributes students' efforts** (requiring students to focus more on skill combinations by a drill-down analysis), compared to Knowledge Tracing models (KT-Single, WKT) and non-hierarchical counterparts. We also showed the benefit of adding external knowledge. Details are reported in [8].

Full paper Author PAWS Lab

[5] González-Brenes, J. P. and Huang, Y. Your model is predictive but is it useful? theoretical and empirical considerations of a new paradigm for adaptive tutoring evaluation. In EDM , pages 187–194, 2015.
[7] Huang, Y., González-Brenes, J. P., Kumar, R., and Brusilovsky, P. A framework for multifaceted evaluation of student models. In EDM, pages 203–210, 2015.
[8] Huang, Y., Guerra, J., and Brusilovsky, P. Modeling Skill Combination Patterns for Deeper Knowledge Tracing. In Workshop on Personalization Approaches in Learning Environments (PALE) in UMAP, 2016.