



Carnegie Mellon University

HCI Prerequisite Relations

Second Progress Report

Team: Elaine Xu, Smeet Poladia, Zhou Lu

Faculty Advisor: Brian Junker

Client: Vincent Aleven

Introduction

- Overview
 - Client: Dr. Vincent Aleven, HCI Institute, CMU
 - In Intelligent Tutoring Systems, data about students, assessments, learning, etc is logged.
 - The project aims to investigate whether prerequisite relations among math topics can be detected in log data.
- Questions to be addressed
 - Test whether topic/skill/unit A is prerequisite for B
 - What metrics of learning and performance?
 - At what level of granularity?

Data

- 3 Workspaces (Grade 7, 500 Students)
 - A = “Analyzing Models of Two-Step Linear Relationships”
 - B = “Modeling Two-Step Expressions”
 - C = “Using Scale Factor”
- Content in A (and presumably KCs in A) are a prerequisite for (presumably KCs in) B
- C is prior to both A & B, and we have no reason to think that C is a pre-req for A or B.
- “MATHia Course 2 (Grade 7) - All Data 2019-2020” (Randomly Selected 500 students, more KCs)

Data

- Take A as an example
 - 43039 rows, 28 columns
 - 500 Students, 7 unique KCs
 - 97 Unique problems, 6 Unique steps

	Anon.Student.Id	Problem.Name	Step.Name	First.Attempt	Incorrects	Hints	Corrects	KC..MATHia.
1	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-5	incorrect	2	0	1	match_slope expression with description.
2	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-6	correct	0	0	2	interpret scenario with numbers
3	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-2	correct	0	0	1	match_indep expression with description.
4	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-4	incorrect	1	0	1	match_linear-term expression with description.
5	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-3	correct	0	0	1	match_dep expression with description.
6	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-1	correct	0	0	1	match_dep expression with description.
7	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-082	gn-classify-item-6	incorrect	2	0	1	match_intercept expression with description.
8	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-082	gn-classify-item-3	correct	0	0	2	match_slope expression with description.
9	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-082	gn-classify-item-1	incorrect	1	0	1	match_indep expression with description.

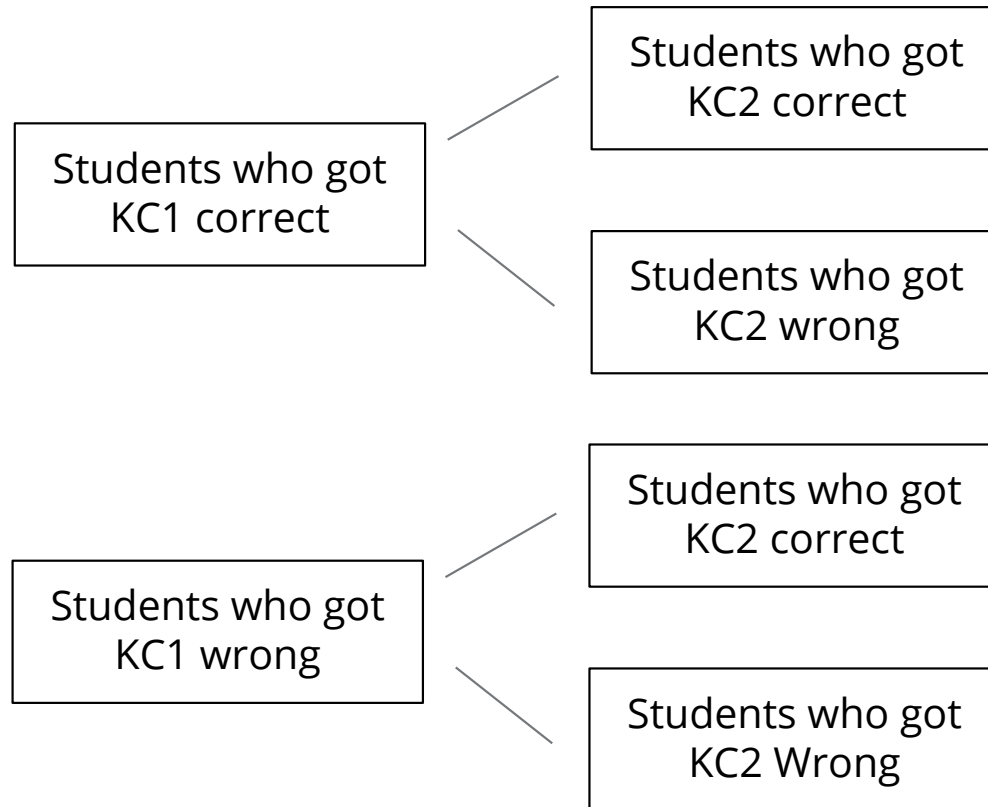
Data

- Knowledge Components
 - interpret scenario with words
 - interpret scenario with numbers
 - match `_slope` expression with description.
 - match `_indep` expression with description.
 - match `_dep` expression with description.
 - match `_intercept` expression with description.
 - match `_linear-term` expression with description.

EDA

- Student's first attempt
 - Correct $\rightarrow 1$
 - otherwise $\rightarrow 0$
- Example:
 - KC1: match_intercept expression with description
 - KC2: match_linear-term expression with description

EDA



$\frac{\text{\#Students who got KC1 correct \& KC2 wrong}}{\text{\#students who got KC1 correct}}$

Error Rate: 49%

$\frac{\text{\#Students who got KC1 wrong \& KC2 wrong}}{\text{\#students who got KC1 wrong}}$

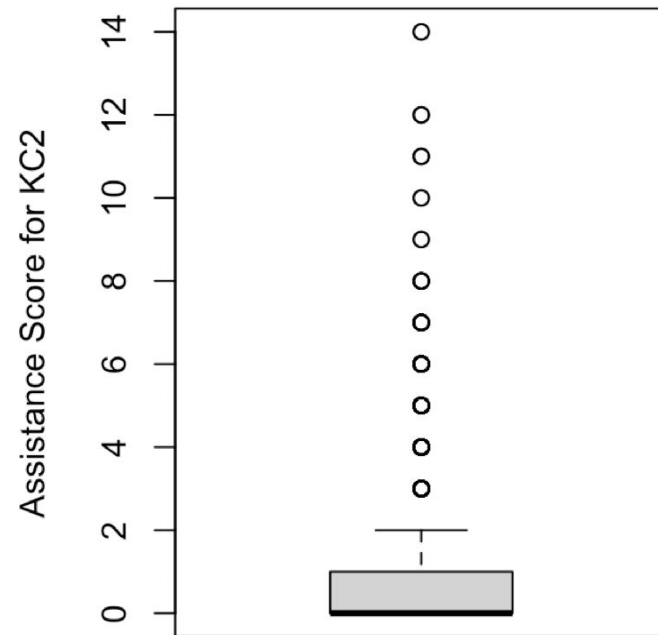
Error Rate: 51%

Method - Assistance Score

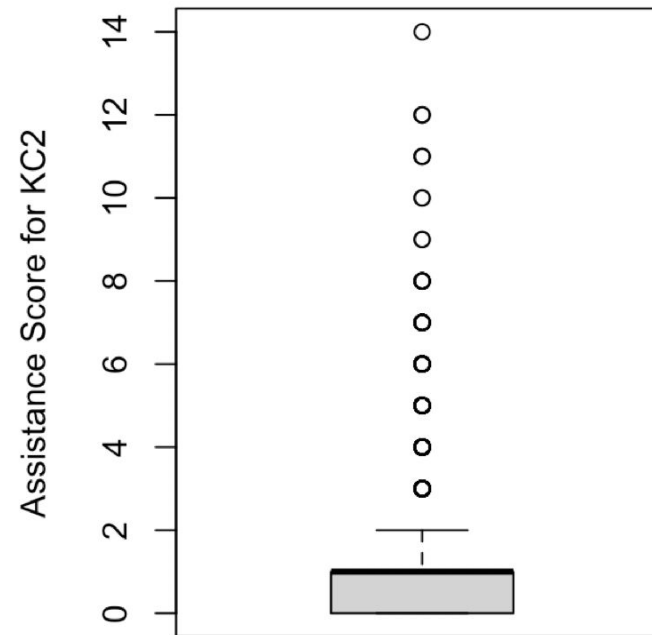
- For a single student,
Assistance Score = **(# of Wrong Attempts + # of Hints requested)** for each step

	Anon.Student.Id	Problem.Name	Step.Name	First.Attempt	Incorrects	Hints	Corrects	KC..MATHia.
1	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-5	incorrect	2	0	1	match _slope expression with description.
2	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-6	correct	0	0	2	interpret scenario with numbers
3	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-2	correct	0	0	1	match _indep expression with description.
4	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-4	incorrect	1	0	1	match _linear-term expression with description.
5	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-3	correct	0	0	1	match _dep expression with description.
6	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-017	gn-classify-item-1	correct	0	0	1	match _dep expression with description.
7	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-082	gn-classify-item-6	incorrect	2	0	1	match _intercept expression with description.
8	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-082	gn-classify-item-3	correct	0	0	2	match _slope expression with description.
9	00e13422-f660-4da6-a068-5f154aa83ae6	analyzing_models_2step_integers-082	gn-classify-item-1	incorrect	1	0	1	match _indep expression with description.

Method - Assistance Score

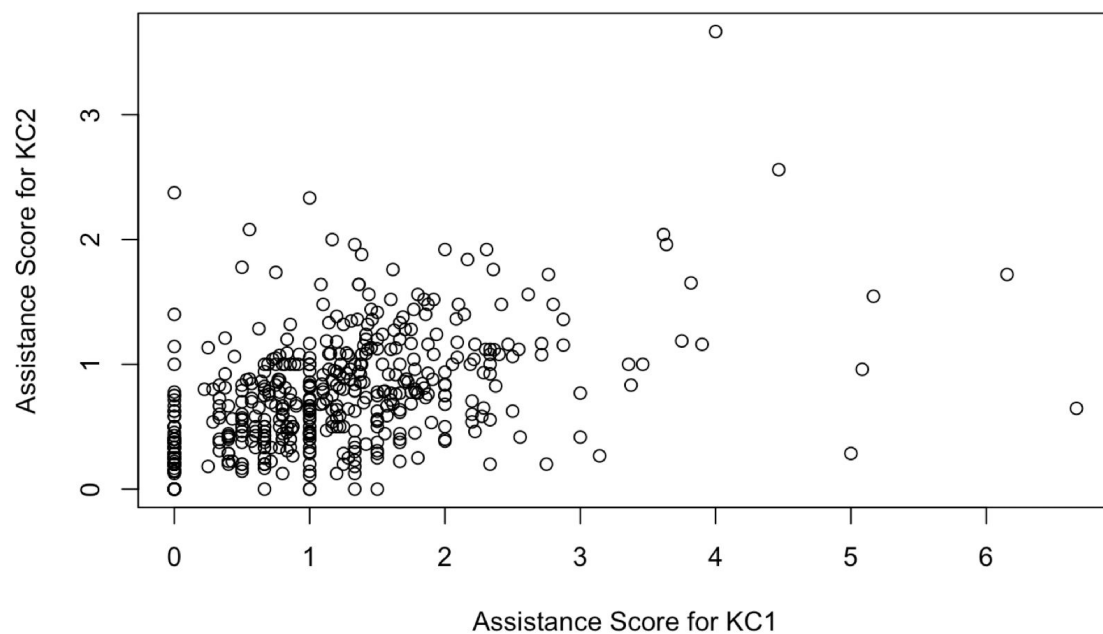


Students who got KC1 correct



Students who got KC1 wrong

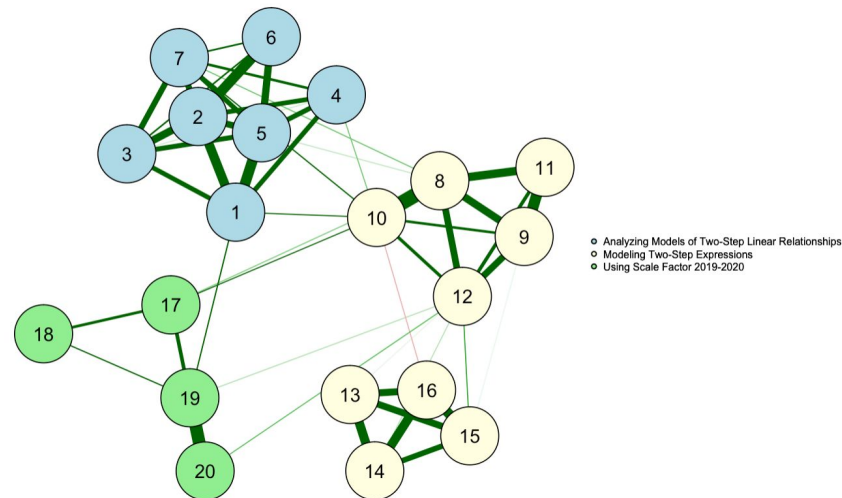
Method - Assistance Score



Correlation between KC1 and KC2:
0.48

Method - Gaussian Graphical Models

- Gaussian graphical models is an exploratory analysis tool that provides relationships between variables in a study.



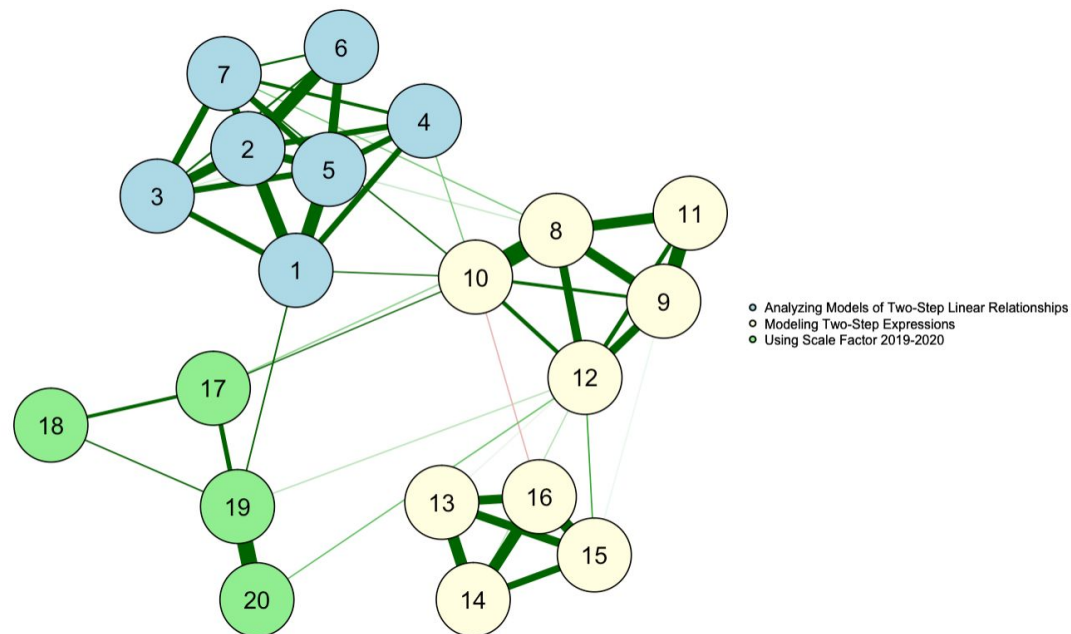
Method - Gaussian Graphical Models

- We have a metric for student performance for all knowledge components.
- Metric we are using currently is Success Rate.

Success Rate = Proportion of correct responses in the first attempt

Anon.Student.Id	match_dep expression with description.	match_linear- term expression with description.	match _intercept expression with description.	interpret scenario with numbers	match_indep expression with description.	interpret scenario with words
00027f1f-dd6d-4b96-96a0-fc815f1862e1	0.6400000	0.5714286	0.6363636	0.3750000	0.6666667	0.6000000
00052b99-8847-4bda-80d0-97cc00c0646b	0.6250000	0.6250000	1.0000000	0.7142857	0.5555556	1.0000000
0005d5fd-cdce-40d6-9b91-2ef847430621	0.3225806	0.4385965	0.4047619	0.5000000	0.5681818	0.5576923
0010b9ba-47f8-4d6b-a69d-31a8ba5dedf7	0.5555556	0.7142857	1.0000000	1.0000000	0.8333333	0.7142857
001d9479-788a-4a1b-aeb6-2be338ff011f	0.6190476	0.6190476	0.7777778	0.5925926	0.7647059	0.8181818

Results - Gaussian Graphical Models

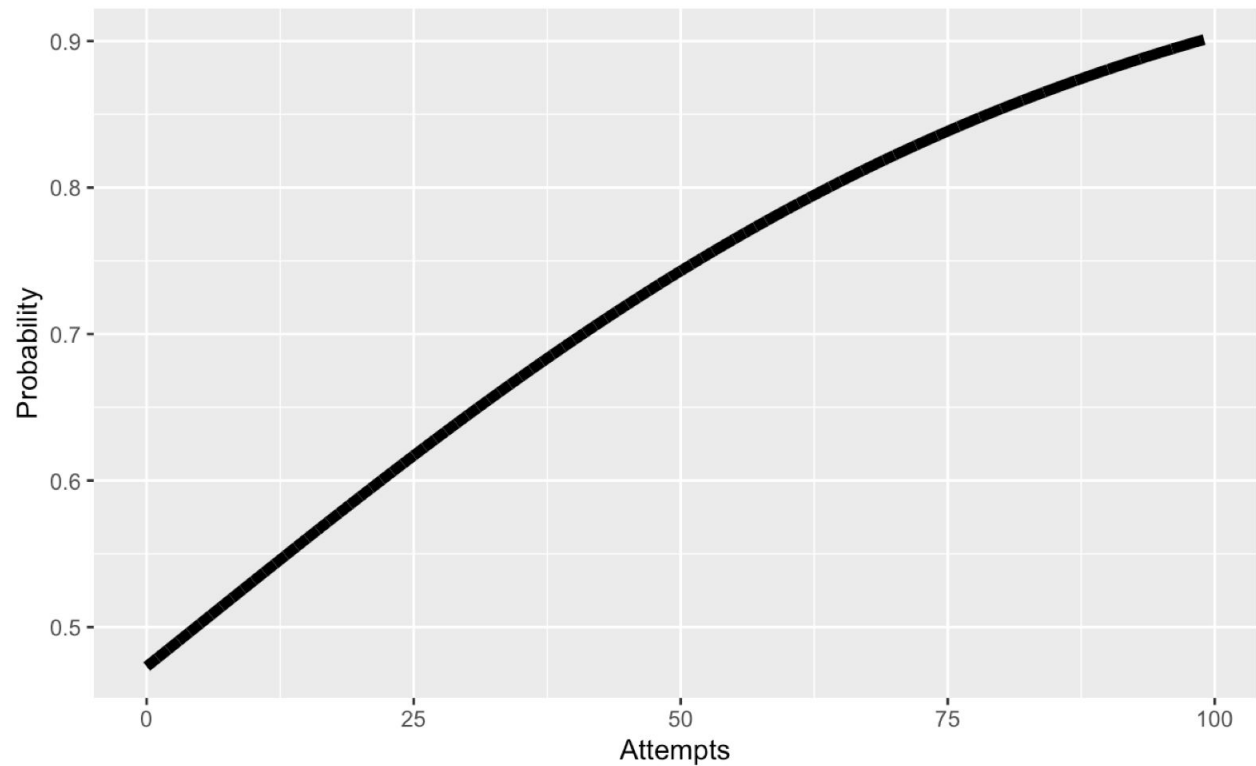


Method & Result - Learning Curve

Logistic Regression

Y: Correct(1), Wrong (0)

X: # attempts



Next Steps

- A metric for student performance and **learning** and examining the relationships between knowledge components again.
- Working on the relationships identified and finding out prerequisite relations.
- Use more predictors in the logistic regression model.

Thank you for your time!