Progress Report

for HCI Learning Discontinuity project

Jie Luo, Naifei Pan, Yiwen Zhang

Agenda

- Introduction
- Our understanding
- Our Approach
- Project Timeline
- Next Steps
- Q&A
- Appendix

Introduction

Advisor:

Zach Branson, Assistant Teaching Professor, CMU

Members:

Name: Yiwen Zhang Background: MSP	Name: Naifei(Julia) Pan Background: MSP	Name: Jie(Jay) Luo Background: MSP

Client Info

Vincent Aleven

- **Professor and Director of Undergraduate Programs in Human-Computer Interaction Institute, CMU**
- **Research Areas**: Enabling Technologies, Learning Sciences and Technologies
- **Co-founder of Carnegie Learning & MathTutor**

Understanding

- This project is tasked with developing a way to detect learning discontinuities within tutor log data to measure effects of out-of-tutor events in Intelligent Tutoring System (ITS)s.
- Issues/Challenges:
  - How helpful are the teacher interventions to students who use ITS?
  - Do these interventions put students on a different learning trajectory, with respect to the specific skills?
  - How can we measure effect?
- Purposes:
  - Improve Learning with ITS
  - Improve scientific understanding of learning with ITS and teachers

Our Approach

Methodology: Potential approaches to consider: AFM model, HSM classifier, Change point analysis, Error rate on subsequent opportunities for the skill, evidence of faster learning following teacher's intervention (Start with small samples)

Our roles: Define the method to evaluate of students' performance and the effect of teacher's intervention. Data analysis & Statistics analysis

Tools and data: R&Datashop, teacher-student-proximity within-tutor-data

Cadence: Group meeting (twice a week, share findings and adjust the research process), meeting with advisor (once a week), meeting with client (once a week)

Deliverables: Statistics analysis report(IMRAD), Code replication file

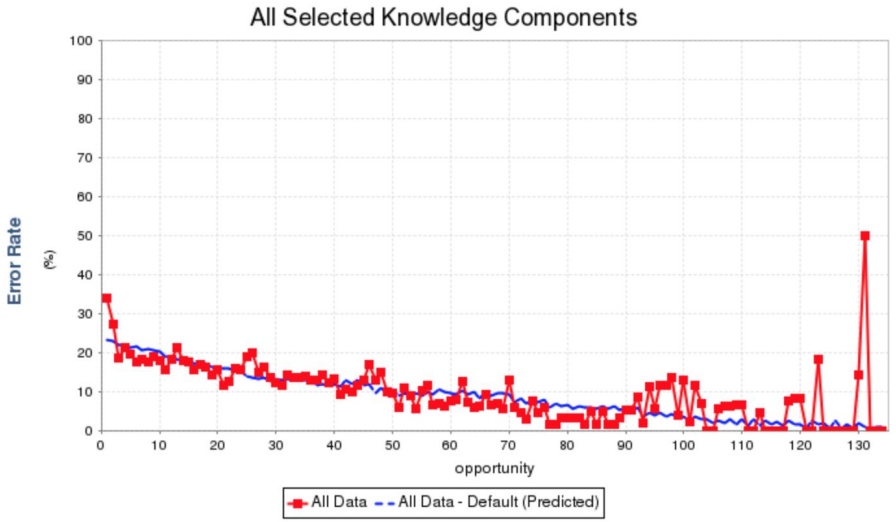
High Level Project Plan (Weekly reports throughout the project)		Feb	Mar	Apr	May
Initiate	Project kick off : - Stakeholder analysis (Internal) - Project charter RACI & Project Plan(Internal) - Status report & charter summary (external)				
Design	Iterative Analysis: - Gather data - literature reviews & define/refine methods - RAID review				
Develop	Iterative Analysis: - EDA & Data cleaning - Model building				
Test	Iterative Analysis: - Performance / Validation				
Finalize	Validate: - Deliverables required by client				
Close	Presentation & Documentation - Summary report & presentation - Additional deliverables				

Next Steps - Data

- Mathtutor/Lumilo
  - source: “Out-of-tutor event detection data” ✓
  - data cleaning ✓
  - simulate fake data ✗
- Carnegie Learning dataset (on hold)
  - 3 files
    - Carnegie Learning LiveHint
    - Carnegie Learning - MATHia 2017-18 Data
    - NGLC Games

Next Steps - Reproduce

- AFM apply to Mathtutor data, generate learning curve



  - other models ✕

Future Steps

- Change-Point Analysis
  - quantify the change: error rate
- *Dashboard Add-ons
  - visualizations that show how learning improve after interventions

Q&A

Appendix: Data

Time	Time.Zone	Duration..sec.	Student.Response.Type	Student.Response.Subtype	Tutor.Response.Type
2013-12-11 14:59:00	America/New_York	4	ATTEMPT		RESULT
2013-12-11 14:59:00	America/New_York	4	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT	tutor-performed	
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT	tutor-performed	
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT	tutor-performed	
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:00:00	America/New_York	3.75	ATTEMPT		RESULT
2013-12-11 15:01:00	America/New_York	3.75	ATTEMPT	tutor-performed	

STATISTICS
57 Columns/ variables
195 Students
2448 Unique Steps
45123 Steps in total

*We used the 'teacher-student-proximity within-tutor-data' dataset accessed via DataShop (pslcdatashop.web.cmu.edu).

Appendix: HSM classifier & AFM model

	Problem ID	Problem-Step	Attempt #	Action	Skill Name	Estimated skill level	HSM result		
2	ANGLE-LINEARPAIR-2	ENTER-SUM	1	ERROR	(ADDITIVE-RELATION)	LOW	Desired Help Seek		
3	ANGLE-LINEARPAIR-2	ENTER-SUM	2	ERROR	(ADDITIVE-RELATION)	LOW	Inappropriate Att		
4	ANGLE-LINEARPAIR-2	ENTER-SUM	3	HINT	(ADDITIVE-RELATION)	LOW	Help Abuse	Desired Help See	33%
5	ANGLE-LINEARPAIR-2	ENTER-SUM	4	HINT	(ADDITIVE-RELATION)	LOW	Help Abuse	Help Abuse:	50%
6	ANGLE-LINEARPAIR-2	ENTER-SUM	5	HINT	(ADDITIVE-RELATION)	LOW	Help Abuse	Inappropriate Att	17%
7	ANGLE-LINEARPAIR-2	ENTER-SUM	6	SUCCESS	(ADDITIVE-RELATION)	LOW	Desired Help Seel		
8	ANGLE-LINEARPAIR-3	ANGLE-1-GIVEN	1	SUCCESS	(LINEAR-PAIR)	MED	Desired Help Seek		
9	ANGLE-LINEARPAIR-3	ANGLE-2-GIVEN	2	SUCCESS	(LINEAR-PAIR)	MED	Desired Help Seek		
10	CLAPPING-PROBLEM-1	ANGLE-2-QUESTIC	1	ERROR	(MULTIPLICATIVE-RELA	LOW	Inappropriate Att		
11	CLAPPING-PROBLEM-1	ANGLE-2-QUESTIC	2	SUCCESS	(MULTIPLICATIVE-RELA	LOW	Desired Help Seek		
12	QUADS-64-1	LENGTH-BLACK-LI	1	ERROR	(ADDITIVE-RELATION)	LOW	Desired Help Seek		
13	QUADS-64-1	LENGTH-BLACK-LI	2	HINT	(ADDITIVE-RELATION)	LOW	Help Abuse		

HSM(help seeking model) classifies each help seeking student's actions as: Desired help, help abuse, and inappropriate attempts

Additive Factors Model (AFM)

Log likelihood that student gets step correct = Student's initial proficiency + Ease of the KC + How much student learned on prior opportunities for this KC

$$\ln \frac{p_{ij}}{1-p_{ij}} = \theta_i + \sum_k \beta_k Q_{kj} + \sum_k Q_{kj}(\gamma_k N_{ik})$$

Given variables

- p_{ij} (0 or 1) probability that student *i* gets step *j* correct
- Q_{kj} (0 or 1) whether KC *k* is needed for step *j* ("*Q matrix*")
- T_{ijk} number of opportunities student *i* has had to practice KC *k*, prior to step *j*

Estimated parameters

- θ_i proficiency of student *i*
- β_k ease of KC *k*
- γ_k gain for each opportunity to practice KC *k*