CS398: Computational Education
• Childhood: Nairobi, Kuala Lumpur
• Ph.D. in AI here at Stanford
• Research lab on Computational Education
Teaching Team
Course mechanics
(this is a light version. Please read the handout for details).
http://web.stanford.edu/class/cs398/
What is CS398?
Graduate Level Class

**CS198**: Teach in small groups

**CS298**: Teach in big groups

**CS398**: Teaching ideas that scale

Enrollment: 30
Learning Goal

Students should be ready to produce a cutting edge project in computational education. Either:

(1) A novel research paper or
(2) A novel learning tool
We have done a good job if you are taking steps towards your potential.
Prereqs

What you really need:

**CS106B (important):**
- Programming

**CS109 (important):**
- Random variables
- Models of random processes
- Coding + probability

**CS 142 or CS 221 (useful)**
- Experience building things
- Experience with AI
Coding in CS398
Staff Contact

- Meet us in class! Ask us questions. Get to know us.
- Go to Office Hours
- Post a question to the slack channel
- Email cs398@cs.stanford.edu
- Email Chris or go to his office for course level issues.
CS398 Units

Hours per week = Units × 3
Average about 8 hours / week for assignments

Start Here

Are you an Undergrad?

Yes

4 Units

No

Do you want to take CS398 for fewer units?

Yes

4 Units

No
Very first offering
The Grand Challenge
People have valued and thought deeply about education for a long time.

John Dewey, 1923
Quality Education Gap

90% of children enroll in primary education [1]
40% in secondary education [1]
20% in tertiary education [1]
Dramatic quality differences

For all learners we want quality.

375 million workers need to be retrained by 2030 [3]
Half a million unfilled computer science jobs (60% of STEM jobs) [2]

Cost Disease of Education

In 2019, there is a unique opportunity to make progress.
Almost perfect...
The Last Remaining Board Game
Self Driving Cars
Computers Making Art
# Smart Phone Access

## Advanced Economies

<table>
<thead>
<tr>
<th>Smartphone</th>
<th>Mobile</th>
<th>No phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>76</td>
<td>17</td>
<td>6</td>
</tr>
</tbody>
</table>

## Emerging Economies

<table>
<thead>
<tr>
<th>Smartphone</th>
<th>Mobile</th>
<th>No phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>45</td>
<td>33</td>
<td>17</td>
</tr>
</tbody>
</table>
3.8 B = 3.8 B

- Poor: 630 M
- Vulnerable: 3.16 B
- Middle Class: 3.59 B
- Rich: 200 M

Source: Projections by World Data Lab
US K12 Students

= 500,000 learners
50,000,000 K12 students in the US

1,234,127 teachers
41,909,399 unique enrolled students
Used in 180+ countries
832M hour of code sessions
Use the power of computers and computation to support learning

=> Open problems in education
=> Open problems in AI
The New York Times

The Year of the MOOC

By LAURA PAPPANO  NOV. 2, 2012
Feedback is Labor Intensive

Online classes have not solved the feedback problem [1].

Computational Education Paradigm

- A student’s past
- Understanding of the student
- Predictions of their future
- Give feedback (hints + summative)
- Visualize knowledge to aid teaching
- Make optimal learning policies
Many domains of student work

Why did the original colonists come to America?
Chapter 1: Deep Knowledge Tracing

*Feedback in a simple context*
Solving for \( y \)-intercept

Solving for \( x \)-intercept

Graphing linear equations

Square roots

Slope of a line

Answer:

Correct

Incorrect
Solving for y-intercept
Solving for x-intercept
Graphing linear equations
Square roots
Slope of a line
What should Riley do next?

Exercise Type:
- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:
- Correct
- Incorrect
Story of Riley

Exercise index

1 10 20 30 40 50

Exercise index

51 60 70 80

Exercise Type:
- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

Answer:
- Correct
- Incorrect

Exercise index
Knowledge Tracing for Feedback

Given $n$ historical answers:

$x_1$, $x_2$, ..., $x_n$

Predict the next one $x_{n+1}$
Build the **first** deep learning algorithm for human learning
Understanding Students

Benchmark AUC

Khan AUC

- Marginal
- BKT
- BKT*
- DKT
Finding the x-intercept
Learns Concept Relationships

Finding the y-intercept
Learns Concept Relationships

Slope of a line
Graphing Linear Equations
Graphing Systems of Equations
Can it help make decisions?
Predicting Next Item

Line graph intuition
Slope of a line
Solving for x-intercept
Solving for y-intercept
Graphing linear equations
Square roots
Line graph intuition
Slope of a line
Solving for x-intercept
Solving for y-intercept
Graphing linear equations
Square roots
Predicting Next Item

Exercise attempted: • correct, • incorrect

Exercise index

Line graph intuition
Solving for y-intercept
Graphing linear equations

We can find what is the “deal maker”
Optimal Teaching

Maximize knowledge after 30 questions

Deep Knowledge Tracing

Previous Algorithms

Blocking

Mixing

Average Predicted Probability

Exercise Index
Chapter 2: Zero Shot Feedback
Online Classes Haven’t Solved the Problem

Feedback is hard
Feedback is hard

Online Classes Haven’t Solved the Problem
The Code.org Dataset

- Students learning **nested loops**
- 50k students with **1.5 million submissions** to a curriculum of **8 exercises**.
- **800 human labels** across 2 of the exercises.
Let's put it all together!

Using your knowledge of for loops and the counter variable, create this drawing where each shape has two more sides than the last. Make sure that each side is 10 times as long as the number of sides in the polygon.

Very little of the code has been provided for you.
Let's put it all together!

Using your knowledge of for loops and the counter variable, create this drawing where each shape has two more sides than the last. Make sure that each side is 10 times as long as the number of sides in the polygon.

Very little of the code has been provided for you.

```
when run
for counter from 3 to 9 count by 2
```

![Shapes progression](image)
Let's put it all together!

Using your knowledge of `for` loops and the `counter` variable, create this drawing where each shape has two more sides than the last. Make sure that each side is 10 times as long as the number of sides in the polygon.

Very little of the code has been provided for you.
Let's put it all together!

Using your knowledge of for loops and the counter variable, create this drawing where each shape has two more sides than the last. Make sure that each side is 10 times as long as the number of sides in the polygon.

Very little of the code has been provided for you.
Instructions

- If there are many moves, focus on the first one.
- Random code strategy is for when the student seems to be trying things randomly.
- Lookout for students who don’t get nesting or pre/post conditions. Often extra blocks in a body is an indication that they don’t get that the post of the loop has to match the precondition.

Question

When run:
- move forward by 100 pixels
- repeat 5 times
- do: turn right by 90 degrees
- move forward by 10 pixels

Label Console

Strategy:
- Beeper Boundary (most people do this)
- Triangle Strategy
- Recursive Strategy

Looping:
- Correct use of looping
- Doesn’t use a while
- Doesn’t have correct stop condition
- Body is missing statements
- Body has extra statements
- Body order is incorrect
- Sets up initial precondition
- Does not get nesting
- Loop post condition doesn’t match precondition
- Repetition of bodies

Cleanup

Record label
It is a very hard problem
Terribly Clever Static Analysis

A student’s program

A bug

Nguyen et al, WWW 2014
// User defined method
private void run() {
  while(isClear()){
    putBeeper();
    move();
  }
  putBeeper();
}

Piech et Al, ICML 2014
method step() {
    putBeeper();
    move();
}
Can we provide feedback by dynamic analysis?

Starter code  First attempt  ...  Final solution
Each node is a unique partial solution.

Pink dots are students.

Each edge is what a teacher suggested.

Solution
The Crowd is Un-wise

Temporal methods tried:
- Shortest path
- Min Time
- Expected Success
- Reinforcement learning
- Most Common Next
- Most Popular Path

- 18%
- 45%
- 12%
SOTA uses RNNs to vectorize programs and classify among K feedback classes.

These models are ...

- Far from human accuracy.
- Uninterpretable... why pick the feedback it did?
- Require lots of labeled examples.

Wang et al, EDM 2017
What Matters for Students?

1. Two compound errors
2. Solves first error
3. Starts reasonable attempt
4. Completes attempt
5. Backtracks
6. Finds solution

Lisa Wang, Chris Piech. Deep Knowledge Tracing on Programming Exercises. L@S 2017
Highly Rates Grit

1. Two compound errors

2. Solves first error

3. Starts reasonable attempt

4. Completes attempt

5. Backtracks

6. Finds solution

Lisa Wang, Chris Piech. Deep Knowledge Tracing on Programming Exercises. L@S 2017
Highly Rates Grit

1. Two compound errors
2. Solves first error
3. Starts reasonable attempt
4. Completes attempt
5. Backtracks
6. Finds solution

Lisa Wang, Chris Piech. Deep Knowledge Tracing on Programming Exercises. L@S 2017
Highly Rates Grit

1. Two compound errors
2. Solves first error
3. Starts reasonable attempt
4. Completes attempt
5. Backtracks
6. Finds solution

Lisa Wang, Chris Piech. Deep Knowledge Tracing on Programming Exercises. L@S 2017
Highly Rates Grit

1. Two compound errors
2. Solves first error
3. Starts reasonable attempt
4. Completes attempt
5. Backtracks
6. Finds solution
Feedback F1 Score

Last Problem (P8)

Label student code

Old Gaurd

Humans

Traditional Deep Learning Doesn’t Work
Label student code

![Graph showing feedback F1 score for Old Guard, Deep Learning, and Humans in the Last Problem (P8).]

- Old Guard
- Deep Learning
- Humans

---

**Inaccurate, Uninterpretable, and Data Hungry**

Piech et al., ICML 2014
Inaccurate, Uninterpretable, and Data Hungry

Label student code

![Diagram showing F1 scores for different methods across problems.](Image)
Label student code

Feedback F1 Score

Last Problem (P8)

- Old Guard
- Deep Learning
- Humans

We need one shot learning

We need verifiability
Four Prototypical Trajectories

[suspense]
Taste of the future of AI
Machine Learning Uses a Lot of Data
Humans Don’t Need Much Data

Single training example:

Test set:
Bayesian Program Learning

A

i) primitives

ii) sub-parts

iii) parts

iv) object template

relation: attached along

relation: attached along

relation: attached at start

type level

token level

v) exemplars

vi) raw data

B

procedure GENERATETYPE

\[ \kappa \leftarrow P(\kappa) \quad \text{Sample number of parts} \]

for \( i = 1 \ldots \kappa \) do

\[ n_i \leftarrow P(n_i | \kappa) \quad \text{Sample number of sub-parts} \]

for \( j = 1 \ldots n_i \) do

\[ s_{ij} \leftarrow P(s_{ij} | s_{i(j-1)}) \quad \text{Sample sub-part sequence} \]

end for

\[ R_c \leftarrow P(R_c | S_1, \ldots, S_{c-1}) \quad \text{Sample relation} \]

end for

\[ \psi \leftarrow \{\kappa, R, S\} \]

return \@GENERATETOKEN(\psi) \quad \text{Return program} \]

procedure GENERATETOKEN(\psi)

for \( i = 1 \ldots \kappa \) do

\[ S_i^{(m)} \leftarrow P(S_i^{(m)} | S_i) \quad \text{Add motor variance} \]

\[ L_i^{(m)} \leftarrow P(L_i^{(m)} | R_i, T_i^{(m)}, \ldots, T_i^{(m)}) \quad \text{Sample part's start location} \]

\[ T_i^{(m)} \leftarrow f(L_i^{(m)}, S_i^{(m)}) \quad \text{Compose a part's trajectory} \]

end for

\[ A^{(m)} \leftarrow P(A^{(m)}) \quad \text{Sample affine transform} \]

\[ I^{(m)} \leftarrow P(I^{(m)} | T^{(m)}, A^{(m)}) \quad \text{Sample image} \]

return \( I^{(m)} \)
1. Enabled teachers to write a student Bayesian Program.

2. We learned to infer into their Bayesian Program (auditable, novel).

- Struggle with double for loops
- Confuses logic for deleting bricks
Generative Understanding

Label student code

Feedback F1 Score

<table>
<thead>
<tr>
<th></th>
<th>Old Guard</th>
<th>Deep Learning</th>
<th>Humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last Problem (P8)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Struggle with double for loops
- Confuses logic for deleting bricks
Label student code

Generative Understanding

Last Problem (P8)

Feedback F1 Score

0.0 0.2 0.4 0.6 0.8 1.0

Old Gaurd Deep Learning Zero Shot Learning Humans

Outstanding student paper award, AAAI 2019

Struggle with double for loops
Confuses logic for deleting bricks
Not just for code
Not just for code

Results from early 2019
Many domains of student work

Why did the original colonists come to America?
Inference is easy in a context free grammar: use A*.

We can highlight parts of the code that caused the model to give a particular feedback.

Replace compiler messages with code highlighting.
Can we grade my CS1 midterm more accurately than Stanford TAs?
Chapter 3: Helping Student Teachers

Presented at SIGCSE 2019
Students should learn the *process* of how to solve programming problems.
But we aren’t providing feedback on *process*.

(We are providing feedback on final solutions)
File: Pyramid.java

/* File: Pyramid.java
 * Name:
 * Section Leader:
 * This file is the starter file for the Pyramid problem.
 * It includes definitions of the constants that match the
 * sample run in the assignment, but you should make sure
 * that changing these values causes the generated display
 * to change accordingly.
 */

import acm.graphics.*;
import acm.program.*;
import java.awt.*;

public class Pyramid extends GraphicsProgram {
    /** Width of each brick in pixels */
    private static final int BRICK_WIDTH = 30;
    /** Height of each brick in pixels */
    private static final int BRICK_HEIGHT = 12;
    /** Number of bricks in the base of the pyramid */
    private static final int BRICKS_IN_BASE = 18;

    public void run() {
        int x = (getWindowSize() / 2) - (BRICKS_IN_BASE / 2) * BRICK_WIDTH;
        int y = getHeight() - BRICK_HEIGHT;
        for (int row = BRICKS_IN_BASE; row > 0; row--){
            sayBricks(x, y, 1);
            BRICK_HEIGHT *= 2;
            x += BRICK_WIDTH / 2;
        }
    }
}

SourceLength

Time Into Problem (hours)

Characters

0 100 200 300 400 500 600 700
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
Selection Comments Code
Pensieve Tool
Students scoring in 99th percentile on midterm exam

Students scoring in ≤3rd percentile on midterm exam

Error Fraction of commits

Fraction of commits

class

Stage 1: single row
Stage 2: nested loop
Stage 3: adjusting nested offset
Stage 4: adding final details

Error
Students work faster and learn more

Hours Spent Coding (minus baseline)

- Baseline
- Exp Term

Pensieve Intervention

7.0 hours

6.3 hours

Midterm Ability

Baseline

Exp Term

Probability Mass

0.0
0.1
0.2

0 5 10

Hours

Homework

Probability Mass

0.0
0.1
0.2

0 5 10
Very Intelligent Tutor in Our Pockets
This will change who can be an amazing teaching.
Do you think that education is **solved**? If not why?
Open Research Problems

Feedback to teachers?
Open Research Problems

Optimal learning ecosystem for refugees
Open Research Problems

Autonomous Tutor Gym
Open Research Problems

Feedback on the process of how someone solves a problem
Open Research Problems

Ad technology for more engaging handouts?
Open Research Problems

Can you translate educational material into different languages?

Can you translate educational material into different languages?
Open Research Problems

Can you learn right student models for **unseen problems**?
Open Research Problems

Can you give feedback to someone engaging in open-ended, unstructured learning?
A Breakthrough for A.I. Technology: Passing an 8th-Grade Science Test

By Cade Metz

Sept. 4, 2019

SAN FRANCISCO — Four years ago, more than 700 computer scientists competed in a contest to build artificial intelligence that
Hi Ayesha, you like skateboarding. If you jump off a half pipe at 2 km/h how high will you go?

What is the relationship between dinosaurs and birds?

Dinosaurs are a diverse group of reptiles of the clade Dinosauria. They first appeared during the Triassic period, between 243 and 233.23 million years ago, although the exact origin and timing of the evolution of dinosaurs is the subject of active research. They became the dominant terrestrial vertebrates after the Triassic–Jurassic extinction event 201 million years ago; their dominance continued through the Jurassic and Cretaceous periods. The fossil record demonstrates that birds are modern feathered dinosaurs, having evolved
Automatic Generation of Contrasting Cases

Like tasting wines side-by-side, contrasts can improve discernment.
Can you measure informal learning?

Greenhouse Gas - USA
Can you measure informal learning?

Greenhouse Gas - USA
Is it learned in school? When?

<table>
<thead>
<tr>
<th>Concept</th>
<th>When its taught (Spring, days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global warming</td>
<td>105</td>
</tr>
<tr>
<td>Fossil fuel</td>
<td>111</td>
</tr>
<tr>
<td>Sea level rise</td>
<td>111</td>
</tr>
<tr>
<td>Greenhouse Effect</td>
<td>114</td>
</tr>
<tr>
<td>Ocean Acidification</td>
<td>118</td>
</tr>
</tbody>
</table>

Bar chart showing popularity of concepts and when they are taught in school.
Where are people learning about Greenhouse Effect
Human Education with AI
What problems do you find interesting?

What is your vision for education's digital future?
Let's dive in
Item Response Theory
How would you solve a problem like this?

Tell me what the students know

Tell me how hard the questions are

What is this is Khan Academy data? What if you get to chose the next item?
What is Item Response Theory

A **model** which predicts if a particular student will answer a particular question correctly
What is Item Response Theory

Used for:
1. Student simulators
2. Reasoning about what a student knows
3. Adaptive tests

A **model** which predicts if a particular student will answer a particular question correctly

“This student has a 70% chance of answering problem 5 correctly”
If student $i$ attempts problem $j$, the likelihood they answer it correctly is...

$$p_{i,j} = \sigma(a_i - d_j)$$
\[ p_{i,j} = \sigma(a_i - d_j) \]

Learner ability: \( a_i \)

Question difficulty: \( d_j \)

\[ \sigma(x) = \frac{1}{1 + e^{-x}} \]

Problem is way too easy for the student

Zone of proximal development

Problem is way too hard

\[ a_i - d_j \]
## Probabilistic Generative Model

<table>
<thead>
<tr>
<th></th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td><img src="circle1.png" alt="Circle" /></td>
<td><img src="circle2.png" alt="Circle" /></td>
<td><img src="circle3.png" alt="Circle" /></td>
<td><img src="circle4.png" alt="Circle" /></td>
</tr>
<tr>
<td>$a_2$</td>
<td><img src="circle5.png" alt="Circle" /></td>
<td><img src="circle6.png" alt="Circle" /></td>
<td><img src="circle7.png" alt="Circle" /></td>
<td><img src="circle8.png" alt="Circle" /></td>
</tr>
<tr>
<td>...</td>
<td><img src="circle9.png" alt="Circle" /></td>
<td><img src="circle10.png" alt="Circle" /></td>
<td><img src="circle11.png" alt="Circle" /></td>
<td><img src="circle12.png" alt="Circle" /></td>
</tr>
<tr>
<td>$a_n$</td>
<td><img src="circle13.png" alt="Circle" /></td>
<td><img src="circle14.png" alt="Circle" /></td>
<td><img src="circle15.png" alt="Circle" /></td>
<td><img src="circle16.png" alt="Circle" /></td>
</tr>
</tbody>
</table>
## Probabilistic Generative Model

<table>
<thead>
<tr>
<th></th>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td><img src="image1" alt="" /></td>
<td><img src="image2" alt="" /></td>
<td><img src="image3" alt="" /></td>
<td><img src="image4" alt="" /></td>
</tr>
<tr>
<td>$a_2$</td>
<td><img src="image5" alt="" /></td>
<td><img src="image6" alt="" /></td>
<td><img src="image7" alt="" /></td>
<td><img src="image8" alt="" /></td>
</tr>
<tr>
<td>...</td>
<td><img src="image9" alt="" /></td>
<td><img src="image10" alt="" /></td>
<td><img src="image11" alt="" /></td>
<td><img src="image12" alt="" /></td>
</tr>
<tr>
<td>$a_n$</td>
<td><img src="image13" alt="" /></td>
<td><img src="image14" alt="" /></td>
<td><img src="image15" alt="" /></td>
<td><img src="image16" alt="" /></td>
</tr>
</tbody>
</table>

**Unobserved variables**
Probabilistic Generative Model

<table>
<thead>
<tr>
<th>$d_1$</th>
<th>$d_2$</th>
<th>$d_3$</th>
<th>$d_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_1$</td>
<td>$\cdot$</td>
<td>$\cdot$</td>
<td>$\cdot$</td>
</tr>
<tr>
<td>$a_2$</td>
<td>$\cdot$</td>
<td>$\cdot$</td>
<td>$\cdot$</td>
</tr>
<tr>
<td>...</td>
<td>$\cdot$</td>
<td>$\cdot$</td>
<td>$\cdot$</td>
</tr>
<tr>
<td>$a_n$</td>
<td>$\cdot$</td>
<td>$\cdot$</td>
<td>$\cdot$</td>
</tr>
</tbody>
</table>

Observed variables
Probabilistic Generative Model

\[ a_i \sim N(\mu = 0, \text{var} = 1) \]

\[ d_j \text{ is given by test-makers} \]

Probabilistic Generative models provide a way to say how likely combinations of values are.

Gives you a method to simulate.

How would you calculate abilities (unobserved) given responses (observed)?
Create an imaginary problem, and an imaginary student and simulate the response
If student $i$ attempts problem $j$, the likelihood they answer it correctly is...

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

$$p_{i,j} = \sigma(a_i - d_j)$$

- Probability correct
- Squashing function
- Ability of student $i$
- Difficulty of problem $j$
Is this the full story?
A brief history of IRT

Charles Darwin writes *Origin of the Species*, 1859

Francis Galton is inspired: creates psychometrics, 1889

Education Testing Services create IRT, 1950

Computers become terribly clever, 2019
If student $i$ attempts problem $j$, the likelihood they answer it correctly is...

\[
\sigma(x) = \frac{1}{1 + e^{-x}}
\]

\[
p_{i,j} = \sigma(a_i - d_j)
\]
If student $i$ attempts problem $j$, the likelihood they answer it correctly is...

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

Probability correct

Guess: probability correct by chance

$$p_{i,j} = c + [1 - c] \cdot \sigma(a_i - d_j)$$

Ability of student $i$

Difficulty of problem $j$
If student $i$ attempts problem $j$, the likelihood they answer it correctly is...

\[ p_{i,j} = c + (1 - c) \cdot \sigma(k_j [a_i - d_j]) \]

- **Guess:** probability correct by chance
- **Discrimination of problem $j$**
- **Probability correct**
- **Ability of student $i$**
- **Difficulty of problem $j$**

\[ \sigma(x) = \frac{1}{1 + e^{-x}} \]
Why simulate students?
Why sigmoid?
Probability Correct

Letter Size (arc mins)

Floored Exponential
Observed
Sigmoid

$c$ (guess prob.)

20/20 20/60 20/100
Floored Exponential

\[ p(x) = \max\left\{ c, 1 - e^{-\lambda_i(x - a_i)} \right\} \]

- Discernibility
- Base ability
- Letter size
- Guess: probability correct by chance
1. Take an eye exam on this website

2. Connect your phone

3. Visualize the math

**Left Eye**

StAT Algorithm

- N done: 15
- MAP acuity: 2.5 arcmin
- Interval: [2.1, 3.6] arcmin

**Likelihood of Acuity Scores:**

- Progress: 75%
Joy of Building
Closest Thing To Magic
Now is the Time
Oh and Its Useful
Everyone is Welcome
Learn By Doing