Education Informatics

Paul Cohen
Carole Beal (LDC and K12@ISI)
Niall Adams (Imperial College, London)
Wes Kerr, Yu-Han Chang, Sinjini Mitra, J-P Steinmetz,
Joshua Moody, Erin Shaw
Education is in trouble

1. The US is losing competitiveness and is vulnerable to technological surprise because we are not training students in science and math. Need to train domestic students.

2. Math and Science are not integrated in most curricula, students don’t see relevance of math or how to apply it to real problems.

3. High proportions of math teachers fail their math teaching certification at least once. U.S. can't replace retiring math teachers quickly enough.

4. Education is the largest budget item of many states, yet our children spend relatively few days in school. Need to accelerate math and science education; e.g., calculus in high school, as in other countries.

5. No US agency addresses all aspects of delivering high-quality K12 education in science and mathematics in one integrated program.

6. In California, one in four students is an English Language Learner. Many other states face similar challenges. Less than 40% pass a State math test at 6th-grade level. Over half of the ELL students enrolled in Algebra 1 courses in Los Angeles schools fail the class at least once; failing Algebra is a strong predictor of dropping out of high school. Of those who graduate, 75% have not completed requirements for entry to state college or university.

7. ....
Educational Data Mining can help

- Modeling utility of AnimalWatch with regression (recent ELL proposal)

**HighELL**: \( \text{PosttestLogOdds} = 0.47 \cdot \text{PretestLogOdds} + 0.28 \cdot \text{Effort} \)

**LowELL**: \( \text{PosttestLogOdds} = 0.45 \cdot \text{PretestLogOdds} + 0.27 \cdot \text{Effort} \)

So INVEST in LowELL students!
Statistical Modeling can help

- Model students as statistical processes
- Estimate parameters of models from many students
- Optimize students' experiences to maximize learning

Examples
- Select the next activity for students working with intelligent tutoring systems
- Model hidden state variables such as engagement and mastery, condition decisions on them
- Segment student sessions into "episodes" and cluster students who have similar episodes (CHUNKing)
- Identify changepoints where students' behaviors change
- Track students' state with nonlinear trackers (e.g., HMMs)
Example: Tutoring Strategies from MDP student models
MDPs and POMDPs for Tutor Actions

\[ Q(s, a) = \lambda \sum_{s' \in S} R(s, a) + \Pr(s'| s, a)V(s') \]

\[ V(s) = \max_a Q(s, a) \]
MDPs and POMDPs for Tutor Actions

For MDPs, guess $R(s,a)$ and learn $\Pr(s'\mid s,a)$ from student performance data to solve for an optimal policy.

$$Q(s,a) = \lambda \sum_{s' \in S} R(s,a) + \Pr(s'\mid s,a)V(s')$$

$$V(s) = \max_a Q(s,a)$$
Example: AnimalWatch data

Top level: starting states along with the frequency of that state in the data in parentheses. 
(-FR 30 345) says that the state with subtract fraction problems with 1 or 2 digits and with 3-5 correct answers in the last 5 occurred 3050 times in the data.

Second level (in box): actions, along with the frequency of that action in the data in parentheses.

The edge between the top 2 levels has the Q-value from the MDP (denoted as “V=…” ) for that (state, action) pair, and the frequency of that pair occurring in the data in parentheses.

Lowest level: transition state s'. The edge between an action and the transition state has the transition probability for the (s,a,s’) triplet denoted by “p =…” and the frequency of that triplet in the data in parentheses.
Example: Rough results

- Control condition: The next problem for a simulated student is selected from the distribution of next problems that we observed in a given state in the AnimalWatch data.

- Experimental condition: The next problem for a simulated student is selected according to the optimal policy.

- Result: In several variants, the MDP condition accrued roughly twice the overall reward for students.

- Caveat: Retrospective, made-up reward function. Prospective study soon.
Education Informatics

- Statistical modeling of students (data mining, modeling, and algorithms) is possible because students work at computers and we can collect data about them
- What else can we do with students who work online?
- Education Informatics is about applications of information technology to education.
World Wide Web and Netizens

DREAMS Communities
- Distributed Science Fair Teams
- Homework clubs
- Peer Tutors
- etc.

Internet Classroom
K12 Integrated Math and Science Curriculum

Common Content Markup
- Topic
- Curriculum
- Media
- Related material

Content Development Tools

Tutoring Tactics and Strategies

Student Models
- Statistical student models of mastery, engagement, ...

Performance Databases
- Easy/hard problems, good/bad paths through curriculum, student clusters, standards, etc

Common Content Delivery Platform

Teachers become managers of individual students' educational trajectories
The Internet Classroom Core Technology

• Content and Data Formats and Markup
• Integrated Math/Science Content
• Common Content Delivery Platform
• Content Authoring Tools
• Assessment and Student Modeling
• Intelligent Tutoring
• Deployment Strategies
Technology Challenge 1
Common Content Delivery Platform (CCDP)

- Today, the best lessons by the best teachers in the US probably aren't deliverable over the web.
- No *common* platform exists for delivering *coordinated* video, images, text, simulations, data sources, analysis tools, drill and test materials, hints and advice, feedback on performance, sources of related material, access to communities of students working on the same material, etc.
- Must work with many kinds of content and variable bandwidth
- Must be controllable by intelligent tutoring systems, teachers, and students
- Must collect performance data at multiple time scales, including sophisticated data collection such as EEGs and eye-trackers
- The challenge: A common *content delivery* platform that provides the student with a common mixed-initiative "look and feel" for all lessons, and provides the content developer with delivery-platform standards to write to, and writes out performance data in standard forms
Technology Challenge 2: Semantic Markup for Content Development, Indexing, Accreditation

• Today, the best material by the best teachers isn't easy to find on the web, hasn't been accredited, doesn't have semantic markup, etc.

• The challenge: Provide content development tools that will
  – Make it easy to author content in many media
  – Link content to extant curricula and find extant material to support the content
  – Submit content for accreditation
  – Ensure that the developed content is compliant with the common content delivery platform
Technology Challenge 3: Assessment

• Today, assessment means taking tests annually. The Internet Classroom makes performance data available from all students on a moment-by-moment basis.

• The challenge: Statistical methods to exploit vast amounts of anonymized information about what our students are doing in the Internet Classroom and how it affects learning. Especially, methods to cluster students for purposes of diagnosis and prediction, and methods that model students over time for purposes of dynamical interventions.

• Examples:
  – Which curriculum units are most/least effective for students like this one?
  – Which paths through a curriculum present fewest pitfalls for such students?
  – What is the student's level of engagement? How is it changing minute-by-minute, day-by-day?
  – Does the student's performance in several curriculum units point to a general deficit/advantage (e.g., does this student have trouble reading, or handling arithmetic?)
Today, there are dozens of intelligent tutoring systems (ITSs),
none used widely. They differ in their approaches to student
modeling, tutoring strategies, target grade level, curriculum, etc.

The challenges:

- Student modeling methods that make accurate predictions about student performance on future problems and curriculum units, with and without interventions from the ITS
- Interventions and tutoring strategies that demonstrably improve students' performance across curriculum units
- The Common Content Delivery Platform must support different tutoring strategies and empirical comparisons of them
Technology Challenge 5: Deployment Strategies

• Today, the classroom is an isolated environment in which the teacher is the nearly sole provider of content for the average student. The Internet Classroom is open to vast amounts of curricular content, supports communities of teachers and students across classrooms and geographic boundaries, and provides teachers unprecedented diagnostic information about each student.

• **The challenge:** Tools to make it easy for teachers to "buy in" to their new role as managers of *individual* students' educations.

• Examples:
  – Fine grained tracking of students' performance, engagement, motivation, within and across curriculum units
  – Tutorials for teachers, themselves, to help them learn how to use Internet Classroom software
  – …
Consequences of Education Informatics

- Teachers and textbook companies are no longer sole providers of content
- The curtain between classroom and home is torn
- Teachers stop talking at the middle of the room and start managing individual students' educations
- Parents might get interested in children's education
- Interested parties, including children, can develop content
- Students get individualized instruction
- The more people who work with these systems, the better their advice to students will be
- Opportunities for ELL students, special needs students,…
Opportunities for K12@ISI

- IES initiatives
- K12@ISI + Games
- Leadership in educational data mining (but RSS has the jump)
- Where the Internet Classroom began (but Andhra Pradesh has the jump)
- Leadership in content/data ontologies and standards
- Clearinghouse for educational technology

- Big opportunities require big investments