Big Data in Online Education

Fethi A. Inan, EdD
Educational Instructional Technology
College of Education
Texas Tech University
Online Education Trends

- Over 6.7 million students taking at least one online course (Allen & Seaman, 2013)
- Majority of doctoral granting universities (80%) offer online courses or programs (Allen & Seaman, 2006)
- At least 15 states already provide some form of virtual schooling (U.S. Department of Education, 2005)
- About 10 percent of all public schools nationwide had students enrolled in distance education courses during 2004-05 (NCES, 2005)
Contributing Factors

- Flexibility (time & place)
- Lower costs
- Life-long learning
- # of degree programs and classes offered
Myths About Online Education

- Online courses are easy
- Online courses are self-paced
- Cheating is more common
- Online learning is suitable for everyone
- The quality of online programs is poor
<table>
<thead>
<tr>
<th></th>
<th>Traditional</th>
<th>Online</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Environment</strong></td>
<td>Classrooms</td>
<td>Anytime, anywhere</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Course management system</td>
</tr>
<tr>
<td><strong>Students</strong></td>
<td>Attend</td>
<td>Participate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Self regulation</td>
</tr>
<tr>
<td><strong>Instructor</strong></td>
<td>Teach (lecture)</td>
<td>Write</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Facilitate (coach)</td>
</tr>
<tr>
<td><strong>Content</strong></td>
<td>Textbooks, Lectures</td>
<td>Electronic resources</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multimedia, Textbooks, Lecture notes</td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td>Synchronous</td>
<td>Asynchronous</td>
</tr>
</tbody>
</table>
Challenges & Issues

- Student diversity
- Student readiness & dropouts
- Student learning and satisfaction
- Faculty professional development
- Faculty workload & support
- Quality of instruction
Competencies for Online Teaching*

*Adapted from Goodyear (2001)
Online Learning 2.0

- Participation
- Communication
- Collaboration
- Community

Learning from Computer
Learning with Computer
Data Mining in Online Courses

Learning Analytics and Educational Data Mining
Applications in Online Learning

- Student skill estimation
- Achievement/performance prediction
- Personalized learning & Intelligent feedback
- Marketing/recruiting
- Behavior detection (engagement)
- Student grouping & collaboration
- Dropout/retention
- Course recommendation
- Social network analysis
- Text Mining
Knowledge Tracing

- **Goal:** Infer student knowledge state from performance (predict knowledge level)
- **Data:** Student performance scores (correct vs incorrect) such as tutorial log that stores student’s answers and time spent per problem
- **Example:** Intelligent tutoring system to individualize practice/test depending on student’s correct/incorrect responses
APS4Math Logic Model

**INSTRUCTIONAL DESIGN**
- Collaborative Development Efforts
- Anytime/Anywhere Access
- Innovative Teaching and Assessment Strategies
  - Individual Diagnostics
  - Targeted Problem Solving
  - Research Based Instructional Principles
  - Modularized Core Curriculum Content

**INDIVIDUALIZED LEARNING**
- Adaptive Assessment
- Adaptive Content
- Adaptive Feedback

**EDUCATIONAL OUTCOMES**
- Improve Student Mastery of Core Math Competencies
  - Student Engagement
  - Problem Solving Skill
  - Cognitive Learning Skills
  - State Test Score
- Enhance Teaching Practices
  - Data Based Decision Making
  - Adaptive Instruction
  - Scaled Use

**IMPACTS**

**EVALUATION & REVISIONS**
Main Question:
The perimeter of a rectangular playing field is 244 feet. If its length is 2 feet longer than twice its width, what are the dimensions of the field?

Steps: 1 of 4

Question:
What unknowns are you looking for?

- The area of the rectangular field
- The length and the width of the rectangular field
- The perimeter of the rectangular field
- The location of the rectangular field

Click the Next Step button
● Total amount of time that students spent studying the tutorial was significantly correlated with the ratio of correctly answered practice questions ($r(57) = .505, p < 0.001$)

● There is significant interaction between beginning prior knowledge groups and knowledge gains from pre to post test, $F(1,53) = 24.162, p < 0.001$

● There was a statistically significant difference between pre and post knowledge test scores ($t(54) = 4.509, p < 0.001$).
Adaptive Learning Systems

- **Goal:** Provide individualized learning opportunities for each student.
- **Data:** Student progress, performance, attitudes, background, and experiences.
- **Example:** Through the incorporation of various instructional strategies, resources, assessments, and interfaces, Adaptive Learning Systems individualize instruction (e.g., content, interface, and strategies) and provide users with more personalized learning experiences.
Probability in Our Daily Lives

Randomness

When an outcome of an experiment is uncertain, it is called a random phenomenon. When you run an experiment and you have a limited number of observations (trials), this is called a short-run. In the short-run, the proportion of times that something happens is highly random, or unpredictable.

Example

If we flip a coin 10 times, we might flip a head seven out of the ten, or 7/10’s of the time. And then on the next 10 flips, we might flip a head two out of ten or 1/5’s of the time. Do you see how the proportion of heads is highly random?

Practice

Simulating the Probability of Flipping a Coin

Click on the Flip Again button at the bottom to simulate the flipping a fair coin.

<table>
<thead>
<tr>
<th>Number of Heads</th>
<th>Number of Flips</th>
<th>Cumulative Proportion</th>
<th>Decimal</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0/1</td>
<td>0.00</td>
</tr>
<tr>
<td>0</td>
<td>2</td>
<td>0/2</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>2/10</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>100</td>
<td>41/100</td>
<td>0.41</td>
</tr>
</tbody>
</table>

Notice how the proportion of heads get closer and closer to 0.5, or ½, as the number of flips gets "larger".
Performance: Adaptive group (M= 6.63) had a significantly higher knowledge gain from the tutorial than non-adaptive group (M= 5.20), [F(1,131)=10.299, p=.002]

Study Time: Students in adaptive group (M=21.17 min) spent more time studying the tutorial than students in the non-adaptive group (M= 18.59 min) [F(1,132)=4.249, p=.041]

Prior Knowledge: Low prior knowledge group benefited more (higher knowledge gain) from the tutorial
Goal: Effective recruiting and marketing

Data: Student demographics & background, enrollment records, degree completion, location of the residence, financial need

Examples: Effective recruiting by predicting potential students’ interest, probability of passing courses and degree completion.
- **Goal**: Prediction of dropout and retention
- **Data**: Student background and performance data
- **Example**: Build early warning systems by profiling potential dropout students using validated models
The location of students’ residence (proximity to the degree offering institution) is related to their decision to dropout/delay.
Goal: Increase student engagement, participation and provide timely feedback and support

Data: Students’ participation, communication, online actions collected by course management systems.

Examples: Grouping related students into categories using their characteristics and participation pattern to make appropriate instructional recommendations and interventions
Student course activities (number of days active, number of chapters/modules completed, number of events/interactions) can be used to predict student completion of the course for a certification.

Students activities in previous courses can predict their successful completion of the current course for a certification.
Fethi A. Inan, EdD
Educational Instructional Technology
Texas Tech University
fethi.inan@ttu.edu

http://www.fethi.org