Educational Data Mining: Potentials for 20th Century Learning Science

Ryan S.J.d. Baker
Teachers College, Columbia University
Exciting Times

• It’s exciting to see this workshop

• Bringing educational data mining research together with all of the other areas represented here

• Another step forward to making EDM one of the key learning sciences of the 21st century
EDM

“Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.”

(www.educationaldatamining.org)
EDM is…

• To Education

• As Bio-informatics is to Biology…

• A new toolbox making possible new kinds of analyses, and analyses at larger scale
EDM is...

• “… escalating the speed of research on many problems in education.”
• “Not only can you look at unique learning trajectories of individuals, but the sophistication of the models of learning goes up enormously.”

• Arthur Graesser, Editor, Journal of Educational Psychology
EDM is…

- “… great.”

- Me
Two communities

• Society for Learning Analytics Research
  – First conference: LAK2011

• International Educational Data Mining Society
  – First event: EDM workshop in 2005 (at AAAI)
  – First conference: EDM2008
  – Publishing JEDM since 2009

• Plus an emerging number of great people working in this area who are (not yet)
Two communities

• Joint goal of exploring the “big data” now available on learners and learning

• To promote
  – New scientific discoveries & to advance learning sciences
  – Better assessment of learners along multiple dimensions
    • Social, cognitive, emotional, meta-cognitive, etc.
    • Individual, group, institutional, etc.
  – Better real-time support for learners
Types of EDM method
(Baker & Yacef, 2009)

• Prediction
  – Classification
  – Regression
  – Density estimation

• Clustering

• Relationship mining
  – Association rule mining
  – Correlation mining
  – Sequential pattern mining
  – Causal data mining

• Distillation of data for human judgment
Types of EDM method
(Baker & Siemens, accepted)

• Prediction
  – Classification
  – Regression
  – Latent Knowledge Estimation

• Structure Discovery
  – Clustering
  – Factor Analysis
  – Domain Structure Discovery
  – Network Analysis

• Relationship mining
  – Association rule mining
Prediction

• Develop a model which can infer a single aspect of the data (predicted variable) from some combination of other aspects of the data (predictor variables)

• Which students are off-task?
• Which students will fail the class?
• Which students will go to college someday?
Structure Discovery

• Find structure and patterns in the data that emerge “naturally”

• No specific target or predictor variable
Structure Discovery

• Different kinds of structure discovery algorithms find…
Structure Discovery

- Different kinds of structure discovery algorithms find… different kinds of structure
  - Clustering: commonalities between data points
    - What are students’ learning strategies in exploratory learning environments?
  - Factor analysis: commonalities between variables
    - Which features of the design of intelligent tutors go together?
  - Domain structure discovery: structural relationships between data points
Relationship Mining

- Discover relationships between variables in a data set with many variables
  - Association rule mining
    - Which course-taking patterns are associated with successful completion of an undergraduate sequence?
  - Correlation and causal data mining
    - Which features of the design of intelligent tutors are associated with better learning?
  - Sequential pattern mining
    - How do patterns of learner hint use over time correspond differently to learning outcomes?
Discovery with Models

• Pre-existing model (developed with EDM prediction methods… or clustering… or knowledge engineering)

• Applied to data and used as a component in another analysis

• Recent review in Hershkovitz et al. (in press)
Distillation of Data for Human Judgment

• Making complex data understandable by humans to leverage their judgment

• Particularly common in the learning analytics community
Types of EDM method

• Many of these overlap with traditional data mining methods

• Some new or special cases that are particularly prominent in EDM
  – Latent Knowledge Estimation
  – Domain Structure Discovery
  – Discovery with Models
Why now?

• Just plain more data available

• Education can start to catch up to research in Physics and Biology...
Why now?

• Just plain more data available

• Education can start to catch up to research in Physics and Biology… from the year 1985
PSLC DataShop
(Koedinger et al, 2008, 2010)

• World’s leading public repository for educational software interaction data
• >200,000 hours of students using educational software within LearnLabs and other settings
• >30 million student actions, responses & annotations
  – Actions: entering an equation, manipulating a vector, typing a phrase, requesting help
  – Responses: error feedback, strategic hints
  – Annotations: correctness, time, skill/concept
The Vision
Can we build educational software

• That knows as much about the learner as e-commerce software knows about its users?
My internet browser cookies know everything about me.
• 35 years old
• Living in New York
• Fluent in English and Portuguese
• Married with children
• College professor
• Reads kindle documents on an ipad
• Travels internationally for work


NEW FLIGHT SEARCH

Walla Walla, WA (ALW) to Ulan Bator, Mongolia (ULN) Fri. Sep. 14 – Thu. Sep. 20, 1 traveler

Show flight summary grid

Earn frequent flyer miles and Expedia Points when you book.
See details

1. Choose Your Departing Flight
Prices are per person and do not include baggage fees or other fees charged directly by the airline.
Displaying UNITED flights | Search for more UNITED flights

Separate Flights | Show Roundtrip Flights | See Flight + Hotel Deals

Sort by: Departure Time | Arrival Time | Stops | Duration | Price

Walla Walla
ALW 6:45am

Ulan Bator
ULN 10:20pm +1 day
2 Stops 24h 35m

Seattle, Seattle
SEA

Roundtrip from $4,296 per person includes taxes & fees

Select

FLIGHT DETAILS
Fri. Sep. 14 (Arriving Sep. 15) Walla Walla, WA to Ulan Bator, Mongolia

Walla Walla (ALW) to Seattle (SEA)
• Stays in cheap hotels
• Likes operas by Philip Glass
• Watches Barney videos online
Used for

- The noble purpose of selling me more stuff
But what do online learning environments know about the learner?
But what do online learning environments know about the learner?

• All too often the answer is...
Some systems model student knowledge

• The systems that incorporate student models typically model what a student knows

• Can be used for
  – Interventions
  – problem selection
  – formative data for teachers
Some systems model student knowledge

• The systems that incorporate student models typically model what a student knows

• Can be used for
  – Interventions
  – problem selection
  – formative data for teachers

• Useful!

• But a one-dimensional picture of the learner
We can…

- Model a lot more than that
My Slogan
My Slogan

Model

Everything!
Because...

• We can probably model everything there is to know about the student...
Because...

• We can probably model everything there is to know about the student...

• More or less, anyways...
In recent years

• There have been proofs of concept – successful models of a range of constructs we might care about

• Can model constructs using just log data

• Some folks also use physical sensors
Physical Sensors

- Facial Recognition Camera
Physical Sensors

- Skin Conductance Sensor
Physical Sensors

- Heart Rate Sensor
Physical Sensors

- Posture “Butt” Sensor
Physical Sensors

• EEG Pendant
Physical Sensors

• FMRI Machine
Physical Sensors

• FMRI Machine

(Coming to a classroom near you!)
Sensors

• Generally lead to better models than just logs
• But not strictly needed
  – People have done surprisingly well at detecting things like emotion just from software logs
Stuff That Someone Has Successfully Modeled and Inferred

(Sensors or no Sensors)
Stuff We Can Infer: Disengaged Behaviors

- Gaming the System (Baker et al., 2004, 2008, 2010; Walonoski & Heffernan, 2006; Beal, Qu, & Lee, 2007)
- Off-Task Behavior (Baker, 2007; Cetintas et al., 2010)
- Carelessness (San Pedro et al., 2011; Hershkovitz et al., 2011)
- WTF Behavior (Wixon et al., UMAP2012)
Stuff We Can Infer: Meta-Cognition

• Self-Efficacy/Uncertainty/Confidence (Litman et al., 2006; McQuiggan, Mott, & Lester, 2008; Arroyo et al., 2009)
• Unscaffolded Self-Explanation (Shih et al., 2008; Baker, Gowda, & Corbett, 2011)
• Help Avoidance (Aleven et al., 2004, 2006)
Stuff We Can Infer: Affect

- Boredom (D’Mello et al., 2008; Sabourin et al., 2011; Baker et al., 2012)
- Frustration (McQuiggan et al., 2007; D’Mello et al., 2008; Sabourin et al., 2011; Baker et al., 2012)
- Confusion (D’Mello et al., 2008; Lee et al., 2011; Sabourin et al., 2011; Baker et al., 2012)
- Engaged Concentration/Flow (D’Mello et al., 2008; Sabourin et al., 2011; Baker et al., 2012)
Stuff We Can Infer: Affect

- Curiosity (Sabourin et al., 2011)
- Excitement (Arroyo et al., 2009)
- Situational Interest (Arroyo et al., 2009)
- Joy (Conati & Maclaren, 2009a, 2009b)
- Distress (Conati & Maclaren, 2009a, 2009b)
- Admiration (Conati & Maclaren, 2009a, 2009b)
- Reproach (Conati & Maclaren, 2009a, 2009b)
Stuff We Can Infer: Deep Learning

• Retention (Jastrzembski et al., 2006; Pavlik et al., 2008; Wang & Beck, 2012)
• Transfer/Shallow Learning (Baker et al., 2011, 2012)
• Preparation for Future Learning (Baker et al., 2011; Hershkovitz et al., under review)
Stuff We Can Infer: Longer-Term Outcomes

- Standardized exam performance (Feng et al., 2009; Pardos et al., in press)

- Dropout (Dekker et al., 2009; Arnold, 2010; Bowers, 2010; Ming & Ming, 2012)
  - Changes in student success in 3rd grade predict who will drop out of high school!

- College attendance (San Pedro et al., under review)
Making Rich Inference about the Student’s Current State

• As a group, what these models show is that we can make rich inference about the student’s current state

• With these models, we can get a multi-dimensional picture of the learner
Making Rich Inference about the Student’s Current State

• As a group, what these models show is that we can make rich inference about the student’s current state, and where they’re going

• With these models, we can get a multi-dimensional picture of the learner
As we model more and more…

• We can envision software that responds to all of these aspects of the student
  – Meta-cognitive adaptation (cf. Arroyo et al., 2007; Roll et al., 2010)
  – Disengagement adaptation (cf. Baker et al., 2006; Walonoski & Heffernan, 2006)
  – Affect adaptation (cf. D’Mello et al., 2010; Arroyo et al., 2011)
  – Instructor interventions when a student is at-risk (cf. Arnold, 2010)
  – Optimizing memory retention (Pavlik et al., 2008; Wang & Beck, 2012)
Supporting…

• Supporting the development of a new generation of more sensitive and effective educational software systems
But also…

- Supporting a leveling-up in the learning sciences
- Enabling richer models of how learning and engagement develop over time
- Enabling multi-dimensional assessment of the effects of learning technologies, and the effects of small and large differences in those technologies
- Enabling quantitative and fine-grained consideration of the effects and inter-
These are exciting times

- It’s great to have all of you along for the adventure

- I’m excited by what we can do together

- And I’m excited to learn from all of you over the next two days
Thanks!

- Follow us on twitter @BakerEDMLab
- Follow us on Facebook Baker EDM Lab
- Publications at http://www.columbia.edu/~rsb2162/publications.html
- "Big Data on Education" on Coursera, Fall 2013
- Concentration in Learning Analytics in TC Masters in Human Development, Fall 2013