Learning Analytics and Educational Data Mining Workshop

New York University – CREATE Lab
April 4–5, 2013

Organizers: Bruce D. Homer*, Jan L. Plass**, & Ryan S.J.d. Baker***

*The Graduate Center, City University of New York
*New York University
*** Teachers College Columbia University

IPoRT
Interdisciplinary Postdoctoral Research and Training Fellowship
FUNDED BY:
ies
THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

Games for Learning Institute

cerego

create

Overview

- Background
- Workshop Goals
- Plan for next two days

Background

- Large Data Sets in Education

Large Data Sets

- Digital technologies in Education and research are producing “Big Data”
  - Extensive Data
  - Intensive Data

Large Data Sets

- Extensive Data
  - Large number of participants
  - Relatively limited number of variables
  - Usually very little demographic information
  - Relatively few observations for each user
  - Wide but shallow data set
- Typical data set for data mining

Large Data Sets

- Intensive Data
  - Relatively low number of participants
  - Large number of observations for each variable
  - Large number of variables for each participant, such as
    - User actions, In-Game Events
    - Survey responses; Extensive demographic information
    - Video Observations
    - Biometric Data (HR, RESP, GSR, EEG, EKG, Pupillometrics)
    - Eye-tracking
  - Narrow but deep data set
- Still developing tools for analyzing: Bayes Nets; ECD

Types of Data

- Operationalize Key Variables based in intensive or extensive data set:
  - **General Trait** Variables (Spatial Ability, Verbal Ability, Executive Functions)
  - **General State** Variables (Prior knowledge, Learning Strategies, Goal Orientation, Self-Regulation)
  - **Situation-Specific State** Variables (Engagement, Emotion, Cognitive Load, Situational Interest)
  - **Learning Behaviors**: (mouse clicks, exploration strategies)
  - **Learning Outcomes**: Skills & Competencies

Data Analyses

- Data Analysis
  - Extensive Large Data Sets: Data Mining techniques
  - Intensive Large Data Sets
    - Parallel streams of real time date (e.g., eye tracking, HR, user logs)
    - Analysis requires detection of patterns across these streams over time

Goals of Workshop

- Develop new and better strategies for collection and analysis of data generated during learning with games & simulations

Intensive & Extensive Data

- How can we use intensive data to enhance extensive data analysis (and vice versa)?
  - Intensive data can help identify the *type* of extensive data being collected
  - Can help identify what patterns to look for in extensive data
  - Validation of observations in extensive data (triangulation)

Workshop Plan

- Today:
  - Introductions
  - Breakout Groups (key topics)
  - Plenary
- Tomorrow:
  - Cross-group idea exchange
  - Reconfigured Breakout groups
  - Plenary
  - Discussion, next steps

Breakout Groups

- Assessment mechanics–design of learning activities that produce meaningful data - Jan Plass, Suzy Letourneau & Bruce Homer
- Analysis of multiple synchronized streams of continuous data - Jay Verkuilen & Jennifer Case
- Data Visualization Techniques for Multiple Data Streams - Claudio Silva
- Data Mining for Behavior Modeling - Ryan Baker