What Do Students Know?

Titus Winters  Tom Payne

University of California, Riverside

ICER Conference, 2005
Outline

Introduction

Process Overview

Data Gathering
  Agar
  HOMR

Data Analysis
  Item Evaluation
  Topic Identification

Conclusions

Questions?
Thoughts

- **Quantitative Assessment and Continuous Improvement**
  - Program assessment is hard – have you seen a good answer?
  - Data mining is a vibrant field
    - But it requires more input data than most courses produce
Thoughts

- Quantitative Assessment and Continuous Improvement
  - Program assessment is hard – have you seen a good answer?
  - Data mining is a vibrant field
    - But it requires more input data than most courses produce
Thoughts

- Quantitative Assessment and Continuous Improvement
  - Program assessment is hard – have you seen a good answer?
- Data mining is a vibrant field
  - But it requires more input data than most courses produce
Thoughts

- Quantitative Assessment and Continuous Improvement
  - Program assessment is hard – have you seen a good answer?
- Data mining is a vibrant field
  - But it requires more input data than most courses produce
UCR’s Plan

Develop a process to

- Enable easy collection of detailed score data
- Require minimum input from instructors
- Provide numeric assessment results
- Make the process easy to use
UCR’s Plan

Develop a process to

- Enable easy collection of detailed score data
- Require minimum input from instructors
- Provide numeric assessment results
- Make the process easy to use
UCR’s Plan

Develop a process to

▶ Enable easy collection of detailed score data
▶ Require minimum input from instructors
▶ Provide numeric assessment results
▶ Make the process easy to use
UCR’s Plan

Develop a process to

▸ Enable easy collection of detailed score data
▸ Require minimum input from instructors
▸ Provide numeric assessment results
▸ Make the process easy to use
Process Overview

- Courses taught as normal, scores recorded at the question level
  - Score matrix $S$ (students $\times$ items)
- Course topics / objectives identified
- Scores analyzed or annotated by instructor for question→topic relevance
  - Relevance matrix $R$ (items $\times$ topics)
- Topics related to curriculum-level objectives – “program outcomes”
  - Course matrix $C$ (topics $\times$ outcomes)
Process Overview

- Courses taught as normal, scores recorded at the question level
  - Score matrix $S$ (students $\times$ items)
- Course topics / objectives identified
- Scores analyzed or annotated by instructor for question→topic relevance
  - Relevance matrix $R$ (items $\times$ topics)
- Topics related to curriculum-level objectives – “program outcomes”
  - Course matrix $C$ (topics $\times$ outcomes)
Process Overview

- Courses taught as normal, scores recorded at the question level
  - Score matrix $S$ (students $\times$ items)
- Course topics / objectives identified
  - Scores analyzed or annotated by instructor for question $\rightarrow$ topic relevance
    - Relevance matrix $R$ (items $\times$ topics)
  - Topics related to curriculum-level objectives – “program outcomes”
    - Course matrix $C$ (topics $\times$ outcomes)
Process Overview

- Courses taught as normal, scores recorded at the question level
  - Score matrix $S$ (students $\times$ items)
- Course topics / objectives identified
- Scores analyzed or annotated by instructor for question $\rightarrow$ topic relevance
  - Relevance matrix $R$ (items $\times$ topics)
- Topics related to curriculum-level objectives – “program outcomes”
  - Course matrix $C$ (topics $\times$ outcomes)
Process Overview

- Courses taught as normal, scores recorded at the question level
  - Score matrix $S$ (students $\times$ items)
- Course topics / objectives identified
- Scores analyzed or annotated by instructor for question→topic relevance
  - Relevance matrix $R$ (items $\times$ topics)
- Topics related to curriculum-level objectives – “program outcomes”
  - Course matrix $C$ (topics $\times$ outcomes)
Process Overview

- Courses taught as normal, scores recorded at the question level
  - Score matrix $S$ (students $\times$ items)
- Course topics / objectives identified
- Scores analyzed or annotated by instructor for question→ topic relevance
  - Relevance matrix $R$ (items $\times$ topics)
- Topics related to curriculum-level objectives – “program outcomes”
  - Course matrix $C$ (topics $\times$ outcomes)
Process Overview

- Courses taught as normal, scores recorded at the question level
  - Score matrix $S$ (students $\times$ items)
- Course topics / objectives identified
- Scores analyzed or annotated by instructor for question→topic relevance
  - Relevance matrix $R$ (items $\times$ topics)
- Topics related to curriculum-level objectives – “program outcomes”
  - Course matrix $C$ (topics $\times$ outcomes)
Process Overview

\[ S \times R \times C = O \]

- \( O \) is students \( \times \) outcomes
- Average columns of \( O \) for numeric outputs
- Course matrix \( C \) done once
- \( S \) requires lots of data
- \( R \) is tedious, but may be automated
Data Gathering

- Two major tools, Agar and HOMR
- Score data stored in spreadsheets, not SQL
  - Minimal format requirements
  - Easy to adopt
  - Gnumeric/ssconvert very useful for automation
  - Integrating with Moodle
Two major tools, Agar and HOMR

Score data stored in spreadsheets, not SQL
- Minimal format requirements
- Easy to adopt
- Gnumeric/ssconvert very useful for automation
- Integrating with Moodle
Data Gathering

- Two major tools, Agar and HOMR
- Score data stored in spreadsheets, not SQL
  - Minimal format requirements
  - Easy to adopt
  - Gnumeric/ssconvert very useful for automation
  - Integrating with Moodle
Agar

Not just another Automated Grader

- CAA - Computer Assisted Assessment
- Rubric creation
Agar

Not just another Automated Grader

- CAA - Computer *Assisted* Assessment
- Rubric creation
Not just another Automated Grader

- CAA - Computer Assisted Assessment
- Rubric creation
Rubric

```
Rubric
Compile
Bulk Test
C++ Driver Test
Correct Filename
Diff Test
Early Late Points
Edit Distance Test
Exit Code
Fatal Error
Invoke Make
On-time Test
Regexp Code
Success

Style
No Global Variables
Good Names
Proper Indentation and Spacing
Line Wraps
Other Style Issues
Good Comments

Front for player names
Front players when it is their turn
Reasonable card layout
Players can "choose" 2 cards
Non-matching choices are flipped back over
Matches removed from board
Game is playable to the end
```

Points

Success Rate

Configure Comments

What Do Students Know?
Not just another Automated Grader

- CAA - Computer *Assisted* Assessment
- Rubric creation
- Write-once comments
Not just another Automated Grader

▶ CAA - Computer *Assisted* Assessment
▶ Rubric creation
▶ Write-once comments
▶ Easy viewing of submissions
▶ Annotation
▶ Automatic mailback of results and feedback
Comments

New Comment

**Comment Name**: Doesn't compile

**Rubric Item**: General Note

**Point Value**: 0

**Description**: Your code doesn't compile at all. As noted in the course syllabus, if your code doesn't compile, we can't grade it, and you get a 0.

**Value Type**: Set value

**Add to current**: Yes

**All rubric items**: Yes

OK

Cancel
Not just another Automated Grader

- CAA - Computer Assisted Assessment
- Rubric creation
- Write-once comments
- Easy viewing of submissions
- Annotation
Not just another Automated Grader

- CAA - Computer Assisted Assessment
- Rubric creation
- Write-once comments
- Easy viewing of submissions
- Annotation
20. Show the 2 trees that result from taking the empty tree and inserting the following values, in order:
10, 20, 30, 40, 50

21. Given a non-empty tree of size 11 and the (not very good) hash function \( h(x) = 2x - 1 \), show the final result of three operations (in order). Use lined paper to make columns.

Add Comment

Comment
Wrong answer, should have been

Points: 2

Rubric Item: Question #02

Current Page 1

Current Submission
Not just another Automated Grader

- CAA - Computer Assisted Assessment
- Rubric creation
- Write-once comments
- Easy viewing of submissions
- Annotation
- Automatic mailback of results and feedback

*Remove redundant work, don’t replace the human.*
Support Existing Paradigms

- Annotation interface for red-pen grading.
- Merge features for breaking up workload among multiple graders
Support Existing Paradigms

- Annotation interface for red-pen grading.
- Merge features for breaking up workload among multiple graders
Homebrew Optical Mark Recognition

- C++ program reads in
  - Page configuration file
  - Number of bubbles to extract
  - Image filename
- Computer vision to normalize page
- AdaBoost classifier to classify each bubble
- Produces “0” or “1” to stdout for each bubble
Homebrew Optical Mark Recognition

- **C++** program reads in
  - Page configuration file
  - Number of bubbles to extract
  - Image filename
- Computer vision to normalize page
- AdaBoost classifier to classify each bubble
- Produces “0” or “1” to stdout for each bubble
Homebrew Optical Mark Recognition

- C++ program reads in
  - Page configuration file
  - Number of bubbles to extract
  - Image filename
- Computer vision to normalize page
- AdaBoost classifier to classify each bubble
- Produces “0” or “1” to stdout for each bubble
Homebrew Optical Mark Recognition

- C++ program reads in
  - Page configuration file
  - Number of bubbles to extract
  - Image filename
- Computer vision to normalize page
- AdaBoost classifier to classify each bubble
- Produces “0” or “1” to stdout for each bubble
Homebrew Optical Mark Recognition

- **C++** program reads in
  - Page configuration file
  - Number of bubbles to extract
  - Image filename
- Computer vision to normalize page
- AdaBoost classifier to classify each bubble
- Produces “0” or “1” to stdout for each bubble
Homebrew Optical Mark Recognition

- C++ program reads in
  - Page configuration file
  - Number of bubbles to extract
  - Image filename

- Computer vision to normalize page

- AdaBoost classifier to classify each bubble

- Produces “0” or “1” to stdout for each bubble
Homebrew Optical Mark Recognition

- C++ program reads in
  - Page configuration file
  - Number of bubbles to extract
  - Image filename
- Computer vision to normalize page
- AdaBoost classifier to classify each bubble
- Produces “0” or “1” to stdout for each bubble
Homebrew Optical Mark Recognition

- C++ program reads in
  - Page configuration file
  - Number of bubbles to extract
  - Image filename
- Computer vision to normalize page
- AdaBoost classifier to classify each bubble
- Produces “0” or “1” to stdout for each bubble
HOMR Features

- Use from any language that can make shell calls
- Highly accurate: 99.99%
- Tolerant: crossed out shapes are “0”
HOMR Features

- Use from any language that can make shell calls
- Highly accurate: 99.99%
- Tolerant: crossed out shapes are “0”
HOMR Features

- Use from any language that can make shell calls
- Highly accurate: 99.99%
  Compare with US Census: 99.6%
- Tolerant: crossed out shapes are “0”
HOMR Features

- Use from any language that can make shell calls
- Highly accurate: 99.99%
  Compare with US Census: 99.6%
- Tolerant: crossed out shapes are “0”
Data Analysis

Two major techniques

- Evaluate individual items
- Determine which items correspond to the same topic.
Data Analysis

Two major techniques

- Evaluate individual items
- Determine which items correspond to the same topic.
Item Evaluation

- Based on IRT
  - Identify difficulty $\beta$ / discrimination $\alpha$
  - Given $\beta$, $\alpha$ is percent correct (below is wrong, above is right).
  - $\beta$ is the split that maximizes $\alpha$

Sample Characteristic Curves
- High Discrimination, Low Difficulty
- Low Discrimination, High Difficulty
Item Evaluation

- Based on IRT
- Identify difficulty $\beta$ / discrimination $\alpha$
  - Given $\beta$, $\alpha$ is percent correct (below is wrong, above is right).
  - $\beta$ is the split that maximizes $\alpha$

Sample Characteristic Curves
- High Discrimination, Low Difficulty
- Low Discrimination, High Difficulty

Titus Winters, Tom Payne

What Do Students Know?
Item Evaluation

- Based on IRT
- Identify difficulty $\beta$ / discrimination $\alpha$
- Given $\beta$, $\alpha$ is percent correct (below is wrong, above is right).
  - $\beta$ is the split that maximizes $\alpha$
Item Evaluation

- Based on IRT
- Identify difficulty $\beta$ / discrimination $\alpha$
- Given $\beta$, $\alpha$ is percent correct (below is wrong, above is right).
- $\beta$ is the split that maximizes $\alpha$
Unsupervised or semi-supervised learning provides many possible techniques:

- Clustering
- Hierarchical clustering
- PCA
- CFA
- NMF
Topic Identification

Unsupervised or semi-supervised learning provides many possible techniques:

- Clustering
- Hierarchical clustering
- PCA
- CFA
- NMF
Unsupervised or semi-supervised learning provides many possible techniques:

- Clustering
- Hierarchical clustering
- PCA
- CFA
- NMF
Unsupervised or semi-supervised learning provides many possible techniques:

- Clustering
- Hierarchical clustering
- PCA
- CFA
- NMF
Unsupervised or semi-supervised learning provides many possible techniques:

- Clustering
- Hierarchical clustering
- PCA
- CFA
- NMF
Unsupervised or semi-supervised learning provides many possible techniques:

- Clustering
- Hierarchical clustering
- PCA
- CFA
- NMF
Results

![Graph showing precision and recall for different methods: NMF, PCA, CFA, and Random. CS008 results are displayed.](chart.png)
Conclusions

- Still room for instructional tool development
- Some analysis is easy
- Some analysis is hard
- Score data is noisy
- Must be mindful of concepts from educational statistics
Conclusions

- Still room for instructional tool development
- Some analysis is easy
- Some analysis is hard
- Score data is noisy
- Must be mindful of concepts from educational statistics
Conclusions

- Still room for instructional tool development
- Some analysis is easy
- Some analysis is hard
- Score data is noisy
- Must be mindful of concepts from educational statistics
Conclusions

- Still room for instructional tool development
- Some analysis is easy
- Some analysis is hard
- Score data is noisy
- Must be mindful of concepts from educational statistics
Conclusions

- Still room for instructional tool development
- Some analysis is easy
- Some analysis is hard
- Score data is noisy
- Must be mindful of concepts from educational statistics
Questions?