Patterns of Behavior in Online Homework

Colin Fredericks
The Point

- Everyone assigns homework.
- Very few studies done on...
  - *How* students do homework,
  - How they do it *best*, and
  - The specific *benefits* of doing it a particular way.
- How much can we learn from homework data?
The Other Point: **Analysis**

- Part of *Researching the Role of Analysis*
- Analysis of problems is a powerful tool
  - Enhances physical intuition
  - Speeds solution
- Does specifically practicing analysis help, or does it only come with time?¹

Data Sources

- Electronic homework at UMass Amherst
- Performance data such as exams and final course grades
- Surveys

- Physics 151, Fall 2003
  - 250 students
  - Eng/Chem/CS
  - 140,000 rows of data

- Physics 181, Fall 2005
  - 55 students
  - Physics/Astro
  - 8,500 rows of data
Methods and Tools

- Excel and Igor
- Correlation Factors
- Principal Component Analysis
Correlation Factors

- Shows degree of linear relationship between two variables
- $r^2$ estimates amount of variance accounted for by a particular variable

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{(\sum X)^2}{N}\right) \left(\sum Y^2 - \frac{(\sum Y)^2}{N}\right)}}$$
What is PCA?

- Part multilinear modeling, part data reduction scheme
- Returns orthogonal vectors that are linear combinations of the original data
- Used in wide variety of fields: chemistry, social science, marketing
- Can be used to group items
- Can be used to identify random data, sort of

Image from https://en.wikipedia.org/wiki/File:GaussianScatterPCA.png
Terminology

Behaviors

Normalized Gauges

Gauges

Raw Data
<table>
<thead>
<tr>
<th>What Data Was Captured?</th>
</tr>
</thead>
<tbody>
<tr>
<td>▸ OWL ID</td>
</tr>
<tr>
<td>▸ Module #</td>
</tr>
<tr>
<td>▸ IU #</td>
</tr>
<tr>
<td>▸ Question #</td>
</tr>
<tr>
<td>▸ Session #</td>
</tr>
<tr>
<td>▸ Attempt #</td>
</tr>
<tr>
<td>▸ Score</td>
</tr>
<tr>
<td>▸ Answer Date</td>
</tr>
<tr>
<td>▸ Answer Time</td>
</tr>
<tr>
<td>▸ Seconds to Respond</td>
</tr>
<tr>
<td>▸ UMass ID</td>
</tr>
<tr>
<td>▸ Question Type</td>
</tr>
<tr>
<td>▸ Due Date</td>
</tr>
<tr>
<td>▸ Due Time</td>
</tr>
</tbody>
</table>

Items in **Boldface** were added later
Preprocessing

- Students with “Incomplete” grade removed
- Course split into engaged vs. disengaged
  - Engaged students attempted 85% of…
    - Homework assignments,
    - Lecture prep assignments,
    - PRS problems,
    - Course feedback surveys, and
    - Quizzes.
    - Attended all exams.
- Trial run indicated viability of study
- Each homework problem categorized
Problem Types

- Analysis
- Conceptual
- Multiple-Choice / Definition
- Traditional
- Problems were categorized by myself and Dr. Leonard
Gauges

- Calculated from raw data
- One specific measurement of student activity
  - Narrowly defined
  - Some seem like duplicates at first
- Usually a count or average
- Literal names:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Seconds to Respond</td>
<td>Number of Attempts</td>
</tr>
<tr>
<td>Time Before Due</td>
<td>Start Time</td>
</tr>
<tr>
<td>Elapsed Time</td>
<td>Short Wrong</td>
</tr>
<tr>
<td>Late Problems</td>
<td>Credit per Attempt</td>
</tr>
</tbody>
</table>
What We Did With Gauges

- Verify validity of problem types
- Correlations with performance
  - Find predictors
  - Understand relations between problem types
  - Compare courses
- Correlations with each other
- Principal Component Analysis
- Combine to form Behaviors
Problem Type Separation

- Gauges along bottom, arbitrary scale on left
- Error bars are twice standard error
- Excellent separation in many cases.
Gauges as Predictors

- **Strong predictors:**
  - Performance-based gauges (but not all)
  - Time-related gauges (but not all)
  - Attempt-related gauges (just about all)
- Courses often differed in gauge correlations.
- Behaviors make better predictors.
Start Time vs Final Grade

\[ R^2 = 0.28968 \]
Late Attempts vs Final Grade

$R^2 = 0.07071$

STR vs Final Grade

$R^2 = 0.12163$

First Attempts vs Final Grade

$R^2 = 0.47153$

SW vs Final Grade

$R^2 = 0.03738$
Gauge Cross-Correlation

- Groups created seem to be more functional than meaningful
- Not as powerful as PCA — no new factors, increased possibility of erroneous correlation

![Table of data](image.png)
PCA on Gauges

- Factors often differ between problem types
- Individual factors, esp. first ones, are often bad predictors. Good pred. come from many factors.
- Combos below: $r=.4$ for exams, $r=.75$ for course
Behaviors

- Linear combinations of Gauges
- Created through intuition and examination
- No “splitting hairs”
- Broader than Gauges
- Evocatively named
## Sample Behaviors

### Uncertainty
- STR: +1
- Attempts: +1
- sw: +1
- Qchange: +1
- Sessions: +1
- Breaks: +1

### Tenacity
- FC: +2
- Fscore: −1
- Problems: +1
- Abandon: −1

### Efficiency
- Attempts: −1
- sw\textsubscript{1}: −0.5
- sw: −0.5
- AvgScore: +1
Other Behaviors

<table>
<thead>
<tr>
<th>Inactivity</th>
<th>Frustration</th>
<th>Slow &amp; Steady</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attempts</td>
<td>Attempts</td>
<td>STR</td>
</tr>
<tr>
<td>sw</td>
<td>sw</td>
<td>Attempts</td>
</tr>
<tr>
<td>swl</td>
<td>Qchange</td>
<td>Elapsed</td>
</tr>
<tr>
<td>Start Time</td>
<td>Sessions</td>
<td>HighAtt</td>
</tr>
<tr>
<td>TBD</td>
<td>Abandon</td>
<td>FC</td>
</tr>
<tr>
<td>Abandon</td>
<td></td>
<td>+1</td>
</tr>
</tbody>
</table>

Grade-Conscious

<table>
<thead>
<tr>
<th>Latt</th>
<th>Fscore</th>
<th>Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>+1</td>
<td>+1</td>
</tr>
</tbody>
</table>

LProb     | +1     |
Tenacity & Efficiency

- Best predictor found
- Also easiest to explain!
  No fancy statistics, better for interpretation than PCA.

<table>
<thead>
<tr>
<th>“r” values</th>
<th>T</th>
<th>E</th>
<th>T+E</th>
</tr>
</thead>
<tbody>
<tr>
<td>P151 Exams</td>
<td>.32</td>
<td>.48</td>
<td>.49</td>
</tr>
<tr>
<td>P151 Course</td>
<td>.76</td>
<td>.32</td>
<td>.80</td>
</tr>
<tr>
<td>P181 Exams</td>
<td>.60</td>
<td>.48</td>
<td>.71</td>
</tr>
<tr>
<td>P181 Course</td>
<td>.81</td>
<td>.37</td>
<td>.83</td>
</tr>
</tbody>
</table>
Behavior and Grade, Physics 181 and 151

\[ R^2 = 0.82268 \]
The Original Point: Analysis

- **181’s Final Exam Question #3**
  - A “How would you solve this problem?” question.
  - Frustration, inactivity, efficiency have no significant impact. Uncertainty is weak positive
  - Grade-conscious, tenacious, slow & steady are best
- “Fruitful Struggle” seems most effective
- Attempting many problems is as useful as getting high scores or starting early.
- Attempting analysis questions more worthwhile, despite higher number of traditional questions
Interesting 151 Survey Items

- Our time-related gauges do not match students’ reporting of time spent
- Disengaged students often wanted to understand material; engaged prefer to improve existing knowledge

- Engaged students more likely to seek multiple resources when stuck
- Disengaged students more likely to give up or “keep trying”
Limitations of Methods & Data

- Behaviorist Bias
- Linear Modeling
- Noise, noise, noise
- Interpretation of higher-order constructs
Wrap-up

- Other approaches appearing in colleges
  - Degree Compass (predicts passing)
  - Course Signals (Nth Week Flag)
- Augmenting grading?
- Data from 8.011 / MITx?
  - Generalize to other disciplines?
  - Other gauges?
  - Longitudinal studies?
Acknowledgements

- William Gerace
- My Thesis Committee
- Ian Beatty
- Emma White

Future Work

- relate.mit.edu
- DontStopLearning.wordpress.com
Other Approaches and Methods (1)

  - Cluster analysis, genetic algorithms, pattern recognition, etc.
  - Optimized results account for ~90% of final grade

- **Warnakulasooriya & Pritchard (2005, MIT)**
  - Using gauges to classify problems by difficulty

- **Kotas & Finck (2002, MSU)**
  - Homework collaboration between students well-correlated with final grade
  - Surveys, log data, and institutional data
Other Approaches and Methods (2)

- **Kortemeyer (2004, MSU)**
  - “Effective Feedback to the Instructor from Online Homework”

- **Cole and Todd (2003)**
  - No significant difference between the performance of students using written or online homework, despite using “multimedia homework with immediate rich feedback.”
  - Suspicions of bleed-through between experimental and control sections: students in pen-and-paper sections sometimes used the logins of the students in electronic homework sections in order to receive feedback.
Research Younger Than Mine

- Lots more “e-homework works” papers, in various fields, especially finance/business.
- Butler, et. al. (2008, Mt. St. Mary’s)
  - “… it was found that the students who received immediate feedback on quizzes had higher quiz and test averages than other students…”
- Kortemeyer (2009, MSU)
  - Gender differences in reported use
- Bennett, et. al. (2007, ????)
  - Data-Mining an Online Homework System