Detecting Changes in Student Behavior from Clickstream Data

Jihyun Park, Kameryn Denaro, Fernando Rodriguez, Padhraic Smyth, Mark Warschauer

University of California, Irvine

NSF Grant No. 1535300
Overview

- Clickstream Data
- Motivation
- Methodology
- Results
- Summary
Overview

CLICKSTREAM DATA
MOTIVATION
METHODOLOGY
RESULTS
SUMMARY
Clickstream Data

- Learning Management System (LMS)
  - New way of interacting with the course material
  - e.g. Canvas, Blackboard, etc.

Clickstream Data from Canvas

<table>
<thead>
<tr>
<th>RandomID</th>
<th>URL</th>
<th>Categories</th>
<th>Action</th>
<th>Created_at</th>
<th>Remote_ip</th>
</tr>
</thead>
<tbody>
<tr>
<td>103445</td>
<td><a href="https://canvas.eee.uci.edu/courses/00">https://canvas.eee.uci.edu/courses/00</a></td>
<td>discussion_topics</td>
<td>index</td>
<td>2016-03-22T19:47:47Z</td>
<td>128.195.96.251</td>
</tr>
<tr>
<td>103445</td>
<td><a href="https://canvas.eee.uci.edu/courses/00">https://canvas.eee.uci.edu/courses/00</a></td>
<td>assignments</td>
<td>show</td>
<td>2016-03-22T19:47:44Z</td>
<td>128.195.96.251</td>
</tr>
<tr>
<td>103445</td>
<td><a href="https://canvas.eee.uci.edu/courses/00">https://canvas.eee.uci.edu/courses/00</a></td>
<td>grades</td>
<td>grade_summary</td>
<td>2016-03-22T19:47:33Z</td>
<td>128.195.96.251</td>
</tr>
<tr>
<td>103445</td>
<td><a href="https://canvas.eee.uci.edu/courses/00">https://canvas.eee.uci.edu/courses/00</a></td>
<td>files</td>
<td>show</td>
<td>2016-03-18T05:38:48Z</td>
<td>128.195.96.251</td>
</tr>
<tr>
<td>103445</td>
<td><a href="https://canvas.eee.uci.edu/courses/00">https://canvas.eee.uci.edu/courses/00</a></td>
<td>modules</td>
<td>index</td>
<td>2016-03-18T05:38:45Z</td>
<td>128.195.96.251</td>
</tr>
</tbody>
</table>

...
Clickstream Data

- Number of click events per day, for each student
  - For example,
Clickstream Data

- Course data from UC Irvine
  - 10-week face-to-face course
  - 85 days
  - 377 students

- Simulated data
  - 85 days
  - 400 students
    - 200 students with rate change at some point
    - 200 students with a single rate
Aggregate Behavior

- Clickstream data from the learning management system (LMS)
Overview

- Clickstream Data
- Motivation
- Methodology
- Results
- Summary
Motivation

- Develop statistical methods
  - More information at the level of individual students
  - Student course engagement

- Change detection technique
  - Detect meaningful change
  - e.g.) Student activity change vs. course outcome
Overview

- Clickstream Data
- Motivation
- Methodology
- Results
- Summary

- Modeling Student Behavior
- Detecting Changepoint
Overview

CLICKSTREAM DATA

MOTIVATION

METHODOLOGY

RESULTS

SUMMARY

Modeling Student Behavior

Detecting Changepoint
Poisson Regression Model

- Poisson distribution
  - Distribution for 'counts'
  - One parameter 'mean': $\lambda$

- Model the mean parameter $\lambda_{i,t}$
  - for student $i$
  - on day $t$

(In our paper we also described Bernoulli model for binary data)
Modeling Student Click Behavior

Function of Population Mean Rate at time $t$

$$\log \lambda_{it} = \mu_t + \alpha_i$$

Individual Random Effect for student $i$

$i$ : student

$t$ : days
Modeling Student Click Behavior

\[ \log \lambda_{it} = \mu_t + \alpha_i \]

- **Pattern**
- **Offset/Shift**

\[ \log \lambda_{it} \]

- **POPULATION**
- **STUDENT1** \( a_i = 0.43 \)
- **STUDENT2** \( a_i = -1.65 \)

\( i : \) student

\( t : \) days

Jihyun Park, Learning Analytics & Knowledge Conference, March 2017
Overview

- Clickstream Data
- Motivation
- Methodology
- Results
- Summary

Modeling Student Behavior
Detecting Changepoint
Changepoint Detection

For each student,

- **Step 1**
  Find a model with changepoint

- **Step 2**
  Select the better model by comparing BIC scores

Compare (M1 Model without Changepoint, M2 Model with Changepoint)
Step 1: Find the Changepoint

- Fit two regression models
  one before the changepoint and one after the changepoint

\[
\log \lambda_{it} = \mu_t + \alpha_{i1}I(t < \tau_i) + \alpha_{i2}I(t > \tau_i)
\]

\(i\) : student
\(t\) : days
Step 1: Find the Changepoint

- Fit **two regression models**
  - one **before** the changepoint and one **after** the changepoint

\[
\log \lambda_{it} = \mu_t + \alpha_{i1} I(t < \tau_i) + \alpha_{i2} I(t > \tau_i)
\]

$i$: student
$t$: days
Step 2: Model Selection

- Is the **Model with changepoint (M2)** better than the **Model without changepoint (M1)**?

- **BIC** (Bayesian Information Criterion)
  - Select the model with **smaller BIC**

\[
BIC = k \log(n) - 2 \log(L)
\]

**Complexity — Goodness of Fit**

- \( L \): Likelihood
- \( k \): Num parameters
- \( n \): Num data points
Step 2: Model Selection, BIC

For each student,

Model without the Changepoint  Model with the Changepoint

BIC_M1 < BIC_M2  • Change

BIC_M1 > BIC_M2  • Change
  • NOT Detected
  • Detected
Result: Simulated Data

![Graph showing simulated data with M1, BIC=135.04 and M2, BIC=98.9. The graph plots changes over time with true changes marked in green, detected changes in red, and raw data in black.](image-url)
Data Sets

- 10-week face-to-face course
  - 85 days
  - 377 students
  - 3 midterms
  - Final exam

- 5-week online course
  - 50 days
  - 176 students
  - 25 video lectures
  - Final exam
Data Sets

- **10-week face-to-face course**
  - 85 days
  - 377 students
  - 3 midterms
  - Final exam

- **5-week online course**
  - 50 days
  - 176 students
  - 25 video lectures
  - Final exam

Results are in the paper!
Preview and Review Behavior

**Total Raw Clicks**
- Very noisy
- 0 ~ 1115 Clicks
- Unimportant information

**File Clicks**
- Lecture notes
- Reading materials
- Exam files (mock-up exams, previous exams, etc.)

**Preview Clicks**
- Opening a file BEFORE the deadline

**Review Clicks**
- Opening a file AFTER the deadline
Preview and Review Behavior

- **Preview (Pre)**
  - Opening a file BEFORE the deadline
  - e.g. opening a lecture note before the lecture starts

- **Review (Post)**
  - Opening a file AFTER the deadline
  - e.g. opening a lecture note after the lecture ends

*Preview results are in the paper*
Example of Two Individual Students

Student **with** detected change

Student **without** detected change
Visualizing Changes in Student Behavior

Increased

Decreased

No-Change
Visualizing Changes in Student Behavior
Student Behavior Change by Grade

- Probability of receiving a passing grade (A, B, C) for Review data
Overview

- CLICKSTREAM DATA
- MOTIVATION
- METHODOLOGY
- RESULTS
- SUMMARY
Summary

- A model for individual student click behavior over time relative to the population of students

- A change-detection method for detecting changes in behavior of individual students

- We were able to
  - Explain individual student's clicking behavior relative to population
  - Relate course engagement with behavior change pattern (increase, decrease)
  - Find relations between behavior change and course outcome
Future Work

- Comparing the same students in multiple courses
- Real-time change detection
- Bayesian models
- Multiple changepoints for longer-term courses

- Python code will be released soon
  - jihyunp@ics.uci.edu
THANK YOU 😊