Quantifying the Mismatch between Course Content and Students’ Dialogue in Online Learning Environments

August 8th, 2017

Sunghoon Lim, Conrad S. Tucker, Kathryn Jablokow, and Bart Pursel

slim@psu.edu, ctucker4@psu.edu, kwl3@psu.edu, bkp10@psu.edu

The Pennsylvania State University
Brick-and-mortar and Online Learning

Classroom

Scalability

Introduction
Feedback and Interactions

Brick-and-mortar learning

VS

Online learning
Introduction

Interactions in Brick-and-mortar Learning

http://www.engr.psu.edu/datalab/
# Interactions in Online Learning

## Introduction

- **What about John Dewey?**
  - Started by Laura · Last post by AE (11 days ago)
  - 0 points · 5 posts · 46 views

- **Gitte's Visual MOOC Takeaway**
  - Started by Brigitte Maronde · Last post by Brigitte Maronde (a day ago)
  - 0 points · 1 post · 11 views

- **MOOC Takeaways**
  - Started by Anonymous · Last post by Mostafa Mohsen Radwan (2 days ago)
  - 8 points · 30 posts · 257 views

- **The juxtaposition of oddities in taking a MOOC**
  - Started by James Robinson · Last post by James Robinson (3 days ago)
  - 3 points · 3 posts · 44 views

---

[http://www.engr.psu.edu/datalab/](http://www.engr.psu.edu/datalab/)
Divergences

Student-discussed topics in discussion forums vs

Topics in course transcriptions
Low and High Divergences

Introduction

- Course transcriptions
- Student feedback data

P(Topic|Document)

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Course transcriptions" /></td>
<td><img src="#" alt="Student feedback data" /></td>
<td><img src="#" alt="Course transcriptions" /></td>
</tr>
</tbody>
</table>

P(Topic|Document)

<table>
<thead>
<tr>
<th>Topic 1</th>
<th>Topic 2</th>
<th>Topic 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="#" alt="Course transcriptions" /></td>
<td><img src="#" alt="Student feedback data" /></td>
<td><img src="#" alt="Course transcriptions" /></td>
</tr>
<tr>
<td>References</td>
<td>Understanding student behavior and feedback</td>
<td>Discovering latent topics from online course content</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>---------------------------------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Outline

1. Textual data extracted from course transcriptions
2. Textual data extracted from student feedback
3. Topic discovery from course transcriptions
4. Topic discovery from student feedback
5. Quantifying the divergences
6. Correlation analysis with the divergences and course performances

LDA
Textual Data from Course Transcriptions

"Concept generation is an integral part of design that begins with..."

Video lectures

Voice to text converters

Textual data
Textual Data from Student Feedback

“I enrolled in the course, thinking it would help me become more creative...”

Online students

Textual feedback from online discussion forums

Method

http://www.engr.psu.edu/datalab/
Latent Dirichlet Allocation (LDA)


http://www.engr.psu.edu/datalab/
Quantifying the Divergences

\( \vec{t}_{1i} \): A topic distribution in course transcriptions for the \( i^{th} \) period

\( \vec{t}_{2i} \): A topic distribution in student feedback data for the \( i^{th} \) period

\[
\text{dis}(t_{1i}, t_{2i}) = 1 - \cos(t_{1i}, t_{2i}) = 1 - \cos(\theta) = 1 - \frac{\vec{t}_{1i} \cdot \vec{t}_{2i}}{|\vec{t}_{1i}| |\vec{t}_{2i}|},
\]

where

\[
\vec{t}_{1i} = \begin{bmatrix}
P(t_1|d_{1i}) \\
P(t_2|d_{1i}) \\
\vdots \\
P(t|T|_1|d_{1i}) \\
P(t_1|d_{2i}) \\
P(t_2|d_{2i}) \\
\vdots \\
P(t|T|_1|d_{2i})
\end{bmatrix},
\]

\[
\vec{t}_{2i} = \begin{bmatrix}
P(t_1|d_{2i}) \\
P(t_2|d_{2i}) \\
\vdots \\
P(t|T|_1|d_{2i})
\end{bmatrix}
\]
Case Study

• A real-world Coursera course (6 weeks)

• Course objective: to provide students with general concepts and principles embedded in a creative problem-solving process

• 1,438 posts and 5,580 comments that are categorized into 6 classes (from Week 1 to Week 6)
Divergences Over Time

![Graph showing divergences over time with different colors representing different numbers of topics.]

-红色：主题数 = 3
-蓝色：主题数 = 5
-绿色：主题数 = 10
-黄色：主题数 = 15
-紫色：主题数 = 20

Case Study

http://www.engr.psu.edu/datalab/
Correlation Analysis (1/4)

\[
r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

- \(n\): Sample size
- \(x_i\): Value of \(i^{th}\) observation from sample \(x\), \(i:1\) to \(n\)
- \(\bar{x}\): Average value of all observations from sample \(x\)
- \(y_i\): Value of \(i^{th}\) observation from sample \(y\), \(i:1\) to \(n\)
- \(\bar{y}\): Average value of all observations from sample \(y\)
Correlation Analysis (3/4)

Correlations: Students' average scores

Students' average scores (0-5) vs. Cosine dissimilarities
Correlation Analysis (4/4)

Correlations: # of submitted assignments

- Correlation Chart showing the relationship between Cosine dissimilarities and # of submitted assignments.
  - The scatter plot displays points with cosine dissimilarities ranging from 0.56 to 0.72.
  - The # of submitted assignments range from 0 to 4,500.
  - A trend line suggests a negative correlation, indicating that as the number of submitted assignments increases, cosine dissimilarities tend to decrease.

Case Study

http://www.engr.psu.edu/datalab/
Conclusion

- The proposed method quantifies the **divergence** between course transcriptions and student feedback data.
- The case study shows that topics in course transcriptions and topics in student feedback data are **mismatched**.
- Cosine dissimilarities **positively correlate** to the total number of posts and comments.
- Cosine dissimilarities **negatively correlate** to the number of submitted assignments and students’ average scores.
Future Work

- Identifying other course performances (e.g., total enrollment, the distribution of academic background) that affect the divergence.

- The effects of the relationship or hierarchy of posts and comments
Thank You.

Acknowledgement

The authors of this work would like to acknowledge the NSF I/UCRC Center for Healthcare Organization Transformation (CHOT), NSF I/UCRC award #1624727 for funding this research. Any opinions, findings, or conclusions found in this paper are those of the authors and do not necessarily reflect the views of the sponsors.


Apaza, Rel Guzman, Elizabeth Vera Cervantes, Laura Cruz Quispe, and José Ochoa Luna. "Online Courses Recommendation based on LDA." In SIMBig, pp. 42-48. 2014.


Liu, I-Fan, Meng Chang Chen, Yeali S. Sun, David Wible, and Chin-Hwa Kuo. "Extending the TAM model to explore the factors that affect Intention to Use an Online Learning Community." *Computers & education* 54, no. 2 (2010): 600-610.


References
