Toward Large-Scale Learning Design:
Categorizing Course Designs in Service of Supporting Learning Outcomes

Daniel Seaton (Harvard University), Dan Davis (TU Delft)

Abstract

This paper applies theory and methodology from the learning design literature to large-scale learning environments through quantitative modeling of the structure and design of Massive Open Online Courses. For two institutions of higher education, we automate the task of encoding pedagogy and learning design principles for 177 courses (which accounted for for nearly 4 million enrollments). Course materials from these MOOCs are parsed and abstracted into sequences of components, such as videos and problems. Our key contributions are (i) describing the parsing and abstraction of courses for quantitative analyses, (ii) the automated categorization of similar course designs, and (iii) the identification of key structural components that show relationships between categories and learning design principles. We employ two methods to categorize similar course designs---one aimed at clustering courses using transition probabilities and another using trajectory mining. We then proceed with an exploratory analysis of relationships between our categorization and learning outcomes.

Introduction

Course structure - the type, frequency, and grading of various activities in a course - has been shown to be a tunable parameter affecting learner behavior and outcomes (Freeman). In traditional on-campus lectures, transitions to active learning formats benefit student behavior, learning outcomes, and attitudes (Laverty, Deslauriers). For example, Laverty et. al changed exam structure from two midterms and a final to bi-weekly exams. Their findings show increased performance, a reduction in guessing and cheating, and improved attitudes (instructors expected students to revolt). There are many other examples that highlight how structure of on-campus courses impacts student outcomes, and it is natural to assume that course structure can be optimized in distance and open-online settings.

By understanding the theory and methodology from the learning design literature and applying it to large-scale learning environments, we see a unique opportunity to bridge the gap between course design decisions and learning behavior and outcomes. The MOOC community has primarily focused on learners so far, but the very nature of open content means that we can freely study design decisions. And because the number of courses accessible to researchers is growing increasingly large (edX has produced roughly 1700 unique courses), there is potential for a new paradigm in studying outcomes through a course design perspective.

In this paper, we attempt to build a framework that can help aid classification of course design in an automated and scalable fashion. Our framework is largely built around the following ideas:

- Parse and abstract courses into quantifiable structural data.

- Measurement of the difference (and/or similarities) between course designs, i.e., clustering courses based on structural data.
- Identification of key structural components that differentiate clusters of courses.

Using a dataset made up of 177 MOOCs from two institutions of higher education, we abstract course design into a sequence of learner activities and apply two types of pattern mining, namely, (i) transition probability mining and (ii) trajectory mining. We explore both methods on an institution by institution basis. In addition, we explore the relationship between our classification (clusters) with a straightforward learning outcome --- verified learner pass rates. This exploratory addition to the study is to further support whether our abstraction and automation can lend itself to goals of improving learning outcomes through better design.

Methodology

Dataset and edX Content

Our dataset consists of edX MOOCs from Delft University of Technology (or DelftX, as it is known on the edX platform) and Harvard University (HarvardX). Within this study, DelftX accounts for 57 MOOCs with a total of 35,283 course components, and HarvardX accounts for 120 MOOCs with a total of 43,514 components. Components are stand-alone assets with which learners interact: videos, problems, html pages, and custom activities. Components are all generally grouped within collections; containers that provide structure and navigation for learners: chapters, sequentials, and verticals.

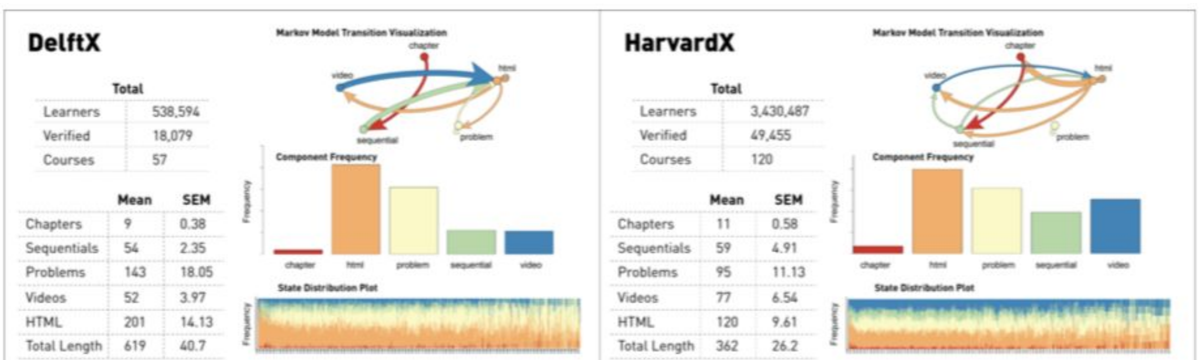

Figure #. Course structure overview for each institution. Tables indicate the total number of enrolled and verified learners for each institution, along with summary statistics about the occurrence of course components (mean per course and standard error of the mean (SEM)). The Markov model transition visualization indicates the most common event type transitions across all courses for each institution; edge/line weights distinguish transition prominence. Component frequency bar graphs show how common each component type was across all courses. The state distribution plot – depicting the left to right occurrence of course components – is a trajectory mining visualization that accounts for the likelihood of component occurrence accounting for all courses in each institution.

All content authored for the edX platform is stored in the Open Learning XML (OLX) format. OLX is a standard that allows the transfer of content between instances of the open source edX platform, authorship outside the platform, and extraction of information related to course design (like in this work). OLX contains the raw markdown (XML) for all authored content in a course, namely, all content tags, text associated with content, and relevant metadata. Courses are generally designed in edX Studio -- a GUI for creating and structuring courses -- masking the OLX from most users. OLX data can be exported through edX Studio and is also provided in regular data exports to edX consortium members through the edX research pipeline. For each course in the present study, we download the OLX data and pass it through a parsing algorithm to structure the data in a more desirable format for analysis (colloquially referred to as the "course axis"). All OLX components are sorted in sequential order according to their placement in the course.

Abstracting Structure From Content

Research in learning design relies heavily on the process of abstracting course structure into a standardized, comparable structure. Abstraction here is the process of stripping away the course topic materials from the underlying structure and components (RQ1). For example, in a course about Statistics, a given sequence of activities might include: a lecture about the difference between frequentist and Bayesian statistics discussion about the benefits and drawbacks of each approach → exam assessing learners' ability to apply what they've learned. The abstracted version of this sequence would become: lecture → discussion → assessment. This method for abstraction is also commonly used when considering learner activity in courses as well.[1,2,3,4] We view this abstraction as similar to processes like coarse-graining in physics, where microscopic structure is often approximated in order to measure macroscopic properties of a system.

Computing Course Similarity

After abstraction of a course, we qualitatively measure the differences between course structures (RQ2) using two approaches: (i) clustering transition probability, and (ii) trajectory mining. Transition probability treats the course activity sequence as a Markov chain and considers the prominence of each of the possible transitions between activity types. The choice for this approach is based on the learning design principle which highlights the importance of the consecutive sequencing of learning activities. The trajectory mining approach takes the entire sequence into account by calculating differences in the order and position of all

---

[1] Boroujeni, Mina Shirvani, and Pierre Dillenbourg. "Discovery and temporal analysis of latent study patterns in MOOC interaction sequences." LAK '18. ACM, 2018.
[2] Davis, Dan, et al. "Gauging MOOC Learners' Adherence to the Designed Learning Path." EDM '16
[3] Pardos, Zachary A., et al. "Enabling Real-Time adaptivity in MOOCs with a personalized Next-Step recommendation framework." L@S '17. ACM, 2017.
[4] Wen, Miaomiao, and Carolyn Penstein Rosé. "Identifying latent study habits by mining learner behavior patterns in massive open online courses." CIKM '14. ACM, 2014.

components, which allows for the analysis of learning design sequences over the span of entire courses beyond single transitions.

Results

Abstracting Structure From Content

In service of RQ1, we find that our abstraction of courses sufficiently enables qualitative insights into course design decisions. HarvardX pivoted toward smaller, modular courses. In some cases, taking long 16 week courses and breaking them up into multiple course---reflected in the average course length. For DelftX, which offers predominantly STEM courses, we confirm a trend towards longer courses containing more assessment activities.

From this method, we find evidence that despite the limited number of elements available in an online learning platform like edX, substantial variation does indeed occur in the learning and structural design of various courses.

Clustering Similar Course Structures

To determine the optimal number of clusters to use with the trajectory mining approach, we again computed clustering quality measures using the Calinski-Harabasz index[5] and silhouette[6] method. We determined the optimal number of clusters for each method and interpreted the results of the clustering by analyzing the defining characteristics of each---uncovering what quantitatively differentiates course designs.

Key Structural Components

With regard to RQ3 which is concerned with identifying the key structural components that define each cluster of similar courses based on quantitative analyses of their syntactic structure, we highlight the qualitative insights offered by each method into the semantic trends which define each cluster. By contextualizing each element into its place in the course relative to other elements, we identify learning design patterns that distinguish each category. For each cluster we find and highlight the key structural components that characterize and differentiate them.

Learning Outcomes

For HarvardX, a one-way ANOVA shows that for the transition probability approach, there is a statistically significant relationship between clusters and completion rates ($p = 0.002$). We therefore conducted a Tukey post-hoc test to identify which pairs of clusters were significantly

[5] Caliński, Tadeusz, and Jerzy Harabasz. "A dendrite method for cluster analysis." Communications in Statistics-theory and Methods 3.1 (1974): 1-27.
[6] Rousseeuw, Peter J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." Journal of computational and applied mathematics 20 (1987): 53-65.

different. We observe significant differences between Clusters 1 and 5 ($p = 0.002$) and Clusters 5 and 6 ($p = 0.004$). The ANOVA model for the trajectory mining approach was not statistically significant ($p = 0.39$). We present any differences strictly as correlation (not causal) and a sign that more work should be done in the future to explore any causality in this relationship. The ANOVA conducted on the DelftX data was not significant.

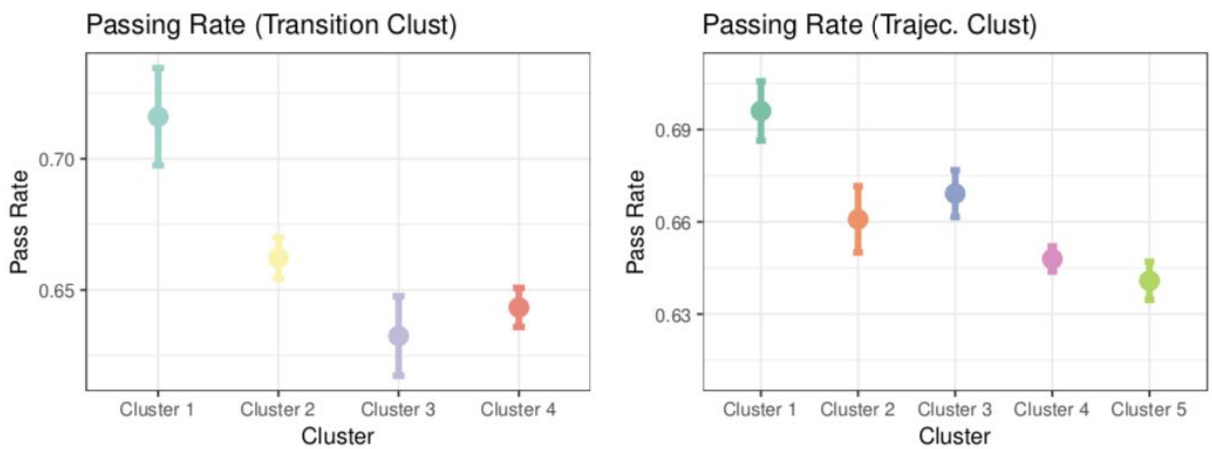

Figure 9. The mean and SEM (error bars) of passing rates of each cluster from DelftX courses.

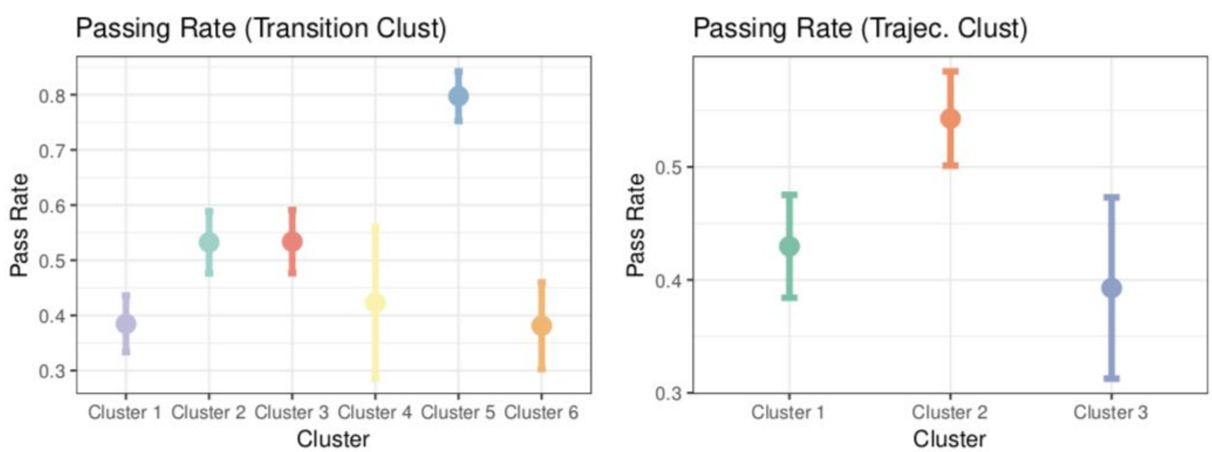

Figure #. The mean and SEM (error bars) of passing rates of each cluster from HarvardX courses.

Conclusion

We are inspired by our ability to automate the process of categorizing course designs and propose that future work needs to continue to refine and test our abstraction method and how it impacts categorization. We also hope to expand our outcome metrics in order to further explore the relationships with course design. Above all, we hope that our work will be a first step in showing the value of addressing digital learning environments from a course structure perspective and finding new challenges as digitization takes an even firmer hold in the learning sciences.