Case Study: A MOOC About MOOCs

I. Introduction

Traditional on-campus education is often prohibitive for people with physical or mental disabilities. The increasing availability of online courses makes it possible for previously marginalized students to attain higher levels of education. The Georgia Institute of Technology's Online Master's of Science in Computer Science (OMS-CS) program is a perfect example. The fully online, blended MOOC format makes postgraduate education possible for those left behind by traditional universities. The low cost (currently $7000 for the program) makes it affordable. The flexible, MOOC-based instruction attracts people who already have enough structured time in their lives. Many students are non-traditional professionals.

In this case study, we describe a project that started with two goals. To present methods and results, both goals involved building a MOOC on an open source platform. After several iterations of researching MOOCs, we chose the Moodle platform and implemented it on Amazon Web Services (AWS). This was a semester-long project for a Computer Science graduate course. The primary functional requirement was to limit costs – AWS' so-called free tier is not always free! Because the project was a prototype, specific considerations customarily needed for a highly scalable MOOC were not a requirement. Proof of concept was the goal.

II. A MOOC about MOOCs

A. *Documenting the Process of Building the MOOC on AWS*

The first part of the project was to build a MOOC about MOOCs. We planned to perform all the Linux systems administration steps to set up and launch the MOOC instance, thus we documented those steps in a MOOC. One video in that MOOC shows how to launch "5-minute Moodle" using the Bitnami Moodle Amazon Machine Image (AMI). This simple approach bypasses hours of Linux command line tasks, yet leaves the student without the knowledge of the Linux command line interface (CLI) environment. Because Moodle is not a "set and forget" application, learners will benefit from the DIY approach in future Moodle administration tasks.

We considered the popular and sophisticated Open edX project but rejected it because it needs more resources than the AWS free tier provides. AWS makes a t2.micro instance possible to run 24/7 for free, but only provides 1Gb of RAM on a single virtual core. We also considered Canvas and Blackboard, but they are expensive, closed source applications.

Moodle is more prevalent in international settings, especially in the developing world. Presumably, this is due to cost. Much research centered on Moodle log data is from Eastern Europe and Latin America. Moodle research papers are written in, and user-bases are also found in, India and broader Asia.

B. *Analyzing MOOC Log Data: Focus on At-Risk Students*

The second part of the project was initially planned to be a new analysis of MOOC log data using machine learning and statistical methods. A common goal in this type of analysis is identifying "at-risk" students who are not likely to pass a course [1,2]. Some learners will sign up for MOOCs never intending to finish all the material. Learners sign up for a wide range of reasons [3] – it is not worth expending effort to retain all of them. The MOOC manager will want to focus on the students who sign up with a full intent to finish. Tracking students manually is an impossible task in a well-subscribed MOOC – instructional staff must apply automated tools, or they will be overwhelmed.

After reviewing available research in this area, it became evident that MOOCs researchers create most of the published papers using data from their own companies or universities. There have been several attempts to share MOOC log data publicly, but the challenges in terms of legal risk and standardizing a schema have been a barrier. This new knowledge led to a different approach where we documented the various ways of sharing MOOC data in a MOOC. We highlighted the successful approaches. This paper tells that story.

III. Problem: A Lack of Anonymized Mooc Data

The multiMOOC Moodle instance successfully applied Agile and Just-in-time Teaching (JITT). We initially planned to obtain an anonymized data set for analysis. Based on older press releases, the Harvard/MIT Person-Course database looked like the right candidate. Once we realized this data set was not available, the search for publicly available data sets began.

One promising repository we looked at was the Pittsburg Science of Learning Center (PSLC) [4]. This is a semi-private repository of education-related data sets. The PSLC is a valuable resource for broader areas of research in Educational Data Mining (EDM) but does not have a focus on MOOCs.

Another project that gave us false hope was Moocdb [5] Moocdb was an initiative with participants such as Stanford University, but Stanford no longer shares data using moocdb. The goal was to create a standardized schema that would allow research that did not need customized algorithms and methods for each specific MOOC's data structure.

MIT and Harvard developed the edx2bigquery open source project [6]. Both universities were exporting their Open edX MOOC data nightly into a data pipeline for offline analysis. At the point where the "nightly" export took longer than 24 hours, they collaborated to create edx2bigquery. The code repo is hosted on Github and open for all to fork. Unfortunately, the data is not available as a public data set in Google's BigQuery.

We also reviewed an attempt to prove that different MOOC schemas could be unified enough to apply ensemble data mining methods in order to predict student dropout rates accurately. The researchers developed an analysis framework for ten edX-based courses at the University of Edinburgh. Having successfully proven the effectiveness of that model, they moved on to a group of Coursera MOOCs. In a section titled, "Transferring Across MOOC Platforms," the authors note that five of their courses shared 21 features, but only 12 with those from Coursera. They developed promising results with the ensemble approach and noted that more research is needed.

In light of these examples, the underlying problem remains. For a university student without a connection to a prominent researcher, that student is not likely to get access to Udacity or Coursera MOOC data. It became apparent that anonymized data sets are the exception, not the rule. The focus shifted. There are now four MOOCs in the Moodle instance:

- metaMOOC – The original "MOOC about MOOCs" teaching how to set up a Moodle instance on AWS. There is a “5-minute Moodle” section plus more than thirty videos showing the Linux systems administration tasks needed to set up Moodle on AWS manually.
- maintMOOC is an ad-hoc MOOC documenting maintenance tasks as they come up. As of this writing, it is AWS specific. As we migrate the site to Google Cloud and link a domain name, we will add these tasks.
- buildMOOC is a MOOC showing how to build out content in a Moodle instance. buildMOOC also has valuable content regarding learning theories and pedagogies. We discuss JITT and Agile there.
- dataMOOC is a MOOC about the various MOOC platforms, highlighting an R-based analysis of a publicly available MOOC log data set from a Moodle.net 2016 Learning Moodle course.

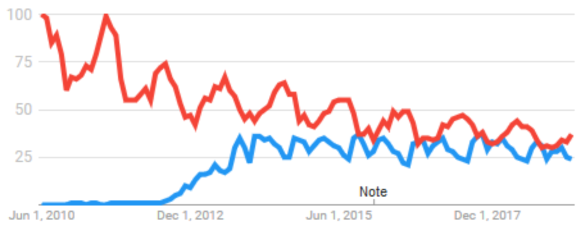
IV. Lessons Learned

A. *Increasing Diversity Through Educational Access*

We learned several valuable lessons in the process of building multiMOOC. As noted in the introduction, MOOCs are leveling the playing field for learners of all socioeconomic backgrounds. People with medical, geographical or financial challenges now have a wide range of choices from free MOOCs to nano-degree programs to blended MOOCs.

Blended MOOCs involve traditional enrollment at a university that teaches some or all of the coursework using online tools. This approach leads to a traditional degree from the institution. In the case of GA Tech’s OMS-CS, it is the same degree and diploma as the on-campus version of the Computer Science program – a program that ranks #8 in the United States!

B. *Expanding the Learner Community Through MOOCs*



An analysis using Google Trends of web-based keyword searches revealed a trend where US-based searches for “MOOC” peaked in 2014 and have trended down, stabilizing only recently. The worldwide trend, shown above, follows a different pattern entirely: searches for “MOOC” (in blue) ramped up until 2013 and have stayed at a plateaued. Worldwide searches also show seasonality, revealing an inverse relationship with searches for “distance education” (in red).” There appears to be an inverse relation with college admission deadlines.

A. *MOOCs and Creative Commons Licensing*

Creative Commons licensing is a boon for people who want to share their work while protecting their rights. In a paper out of the MIT Teaching Systems Lab, the authors note that traditional MOOCs are “open” in more ways than one.

MOOCs have open registration without cost to anyone with an email address. The content is “open,” meaning that anyone who registers can access it [7]. Another form of openness that the multiMOOC courses have adopted is open licensing. We currently have chosen the Attribution-NoDerivatives 4.0 International license that allows users a great deal of flexibility in reusing our content, as long as there is proper attribution.

V. Conclusion

The multiMOOC project was a success by several measures. We learned the landscape of MOOC log research. We identified areas that need improvement. Sharing data openly has not been realized. Having publicly available anonymized MOOC data sets will lead to a better understanding of how to improve low completion rates. Educational outcomes will likely improve if students know they are investing their time and resources well.

More importantly, we were able to apply and synthesize previously held skillsets from the world of Agile, Lean Management, and Just-in-Time business strategies. A newfound interest in learning theories was spurred on by the second author, who was also our Mentor during the project. Without her and Dr. Joyner, this would not have been possible. (We also thank Filipe Altoe for editing and formatting assistance.)

References

[1] N. Payne, D. Hegberg, and D. Joyner, “Identifying the Factors That Predict Academic Performance within a Graduate Massive Open Online Course (MOOC),” 2015.

[2] D. Joyner, “Toward CS1 at scale: building and testing a MOOC-for-credit candidate,” in *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, 2018, p. 59.

[3] R. W. Crues, N. Bosch, C. J. Anderson, M. Perry, S. Bhat, and N. Shaik, “Who they are and what they want: Understanding the reasons for MOOC enrollment.”

[4] R. K. Sawyer, *The Cambridge handbook of the learning sciences*. Cambridge University Press, 2005.

[5] K. Veeramachaneni, S. Halawa, F. Dernoncourt, U.-M. O’Reilly, C. Taylor, and C. Do, “Moocdb: Developing standards and systems to support MOOC data science,” *arXiv Prepr. arXiv1406.2015*, 2014.

[6] G. Lopez, D. T. Seaton, A. Ang, D. Tingley, and I. Chuang, “Google BigQuery for Education: Framework for Parsing and Analyzing edX MOOC Data,” in *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, 2017, pp. 181–184.

[7] E. Huttner-Loan, G. Beazley, C. Glenwerks, A. Napier, J. Littenberg-Tobias, and J. Reich, “Making a Creative Commons MOOC: Challenges and Opportunities,” in *2018 Learning With MOOCS (LWMOOCS)*, 2018, pp. 81–84.