metaMOOC: A MOOC About MOOCs

James Lohse
CS 6460 GA Tech
jlohse3 – Project Proposal
jiml@gatech.edu

INTRODUCTION

This project, metaMOOC, is primarily on the content track but involves aspects of a development track project. The goal is an Internet-available Moodle-based Massively Open Online Course (MOOC) that teaches Education Data Mining (EDM) and Learning Analytics (LA) researchers where to find public data sets and how to setup a Moodle-based MOOC.

These researchers will be attempting MOOC analyses predicting academic performance with an eye toward early identification of at-risk students. By identifying at-risk students (those who have a serious intent but are likely not to succeed), MOOC instructors can allocate resources more efficiently in ways that increase retention rates, a known problem with MOOC courses.

The metaMOOC Moodle instance will be installed on an Amazon Web Services (AWS) Free Tier account. It requires PHP 7.0, mysql and an Apache2 http server. Ubuntu 18.04 will be used.

The full project will involve development of the MOOC in three or four phases:

> 1) Set up a free tier AWS account and configure a self-hosted Moodle instance at the system level using Ubuntu 18.04
> 2) Develop content that documents and teaches the steps needed for part 1

3) Create content that teaches how to understand publicly available data sets from other MOOC platforms
4) As time allows, go more in-depth into issues such as the Linux commands needed for Moodle, discuss whether Moodle can be deployed on Mac or Windows machines, looking at other open MOOC platforms (edX, Udemy), debugging Moodle self-hosted instances, and MoodleCloud.

WHY IS THE PROPOSED PROJECT VALUABLE?

Each day, hundreds of thousands of users interact with MOOCs offered by edX, Coursera and Udacity. Currently, there are many research papers available that study that data. The studies typically aim to identify at-risk students to improve retention rates and student success.

On the surface, it appears that anyone with a data science background can do this type of study. It doesn't take long to realize that most all studies are done by the same organization that publishes the MOOC they are studying. In those cases, researchers have access to proprietary data and can safely ignore de-identification. As long as their study results don't include any personally identifiable information (PII), they don't take any risks associated with re-identification.

If a researcher does not work within an organization that runs the MOOC generating the data, several ethical and technical complexities come into play. First, the process of de-identifying (anonymizing) data can corrupt the relationships in the data and introduce a technical form of bias into the resulting research. Second, institutions that produce the data have legal liability for any such misuse or re-identification of PII, so they tend to avoid this risk. Another reason to avoid sharing MOOC data publicly is simply the time and effort needed to clean the data of PII. Finally, there are a variety of data formats, even for different MOOCs within the same institution.

Fortunately there are a few sources to obtain publicly available de-identified data, but there is a learning curve to understanding the different formats.

WHO MAY BENEFIT FROM THIS PROJECT?

MOOCs have fallen short of their goals. In 2012, Stanford Professor Sebastian Thrun is credited with creating the MOOC when he posted lecture videos of an Artificial Intelligence course online for anyone to view. In those heady days, the hope was that MOOCs would level the educational playing field, making high quality education

available to at-risk populations in less developed countries who are deprived of local opportunities for a technical education.

It wasn't long before MOOC producers realized that their courses were largely reaching the already tech-literate populations in highly developed countries. They also realized that the completion rates and retention rates were dismal.

The benefits of this project are two-fold. Researchers who do analysis on high-integrity publicly available data sets can share research results in a way that is reproducible, thus furthering the scientific validity of the studies. The more researchers learn about reaching at-risk populations, the more those populations benefit.

RELATED WORK

There are a few blog posts that explain specific data sources, such as Lopez' article about Google BigQuery and the edx2bigquery repo. The Pittsburgh Science of Learning Center (PSLC) DataShop offers a wide variety of education-related data sets but little related to MOOCs. It can be difficult to separate the wheat from the chaff on the DataShop when it comes to MOOC data. Stanford previously shared data using the moocdb standard, but stopped sharing in 2018.

To the best of this author's knowledge, there is an unmet need to provide a higher level view of these disparate data sets. There is a need to package this knowledge into a "meta-MOOC" such that researchers can register for the MOOC, review selected content that they find useful, and then apply that knowledge. metaMOOC will provide a road map to these data sets where no road map exists today.

metaMOOC also will include a section showing how to deploy a popular open source MOOC hosting system, Moodle. In fact, metaMOOC will be deployed on Moodle! The lessons learned by this author in deploying the Moodle instance will be documented and posted in metaMOOC.

To this author's knowledge, the only MOOC-based sources for this knowledge are:

1) A MOOC presented by LearnMoodle (https://learn.moodle.net/), but the course is not available until June, 2019. Currently the LearnMoodle MOOC is not accessible and new account cannot be created. By getting metaMOOC in beta by March and finished by May, it's possible to capture some of the interest and search traffic ahead of the June course. In addition, this course teaches different subject matter than this project.

2) Lynda.com has a four year old lesson showing how to install Moodle. It's for much older versions of the software: https://www.lynda.com/Moodle-tutorials/Moodle-Administration-Essential-Training/372439-2.html. This lesson does not cover using an AWS cloud instance.

3) There is also a Lynda.com lesson on MoodleCloud, which is different from the self-hosted AWS instance covered by this project (https://www.lynda.com/Moodle-tutorials/MoodleCloud-First-Look/419131-2.html). In MoodleCloud, for the free tier, there is a limit of fifty students and inactive sites (sites not accessed for sixty days) are deleted.

metaMOOC will initially allow any registered user to view the content at any time and in any order. It will be the only place to review such information at the learner's leisure, unlike the scheduled MOOC offered in the first item above.

NOTE: While metaMOOC will not have hard limits on the number of enrolled students, the type of AWS instance will limit CPU resources and therefore will limit the number of concurrent users.

PROPOSED WORK

This section assumes a reader with the following skills could implement the project based on the instructions:

1) Knowledge of Linux command line and systems administration (Ubuntu 18.04)
2) Understanding how to install and configure supporting packages including: Apache2 web server,
3) Understanding the AWS system well enough to establish SSH access from a local machine, deploy a public-facing web server, understand networking and security concepts at the minimum level required by AWS
4) Write scripts, record video, integrate video and written content into the Moodle instance
5) The soft skill of being capable of debugging problems and configurations in a Linux environments

Tools Needed

metaMOOC will be built within a Moodle 3.1 installation. Moodle requires PHP 7.0, an Apache2 http server and mysql database

metaMOOC will be deployed on a Free Tier AWS instance of Ubuntu 18.04, so an AWS account must be established

AWS components needed include Ubuntu instance, S3 storage, default security group, single subnet with public endpoint, Elastic IP associated with instance and a domain name that points to the Elastic IP address for easier public access.

Content to be Produced

metaMOOC will consist of 3-4 sections containing video tutorials. OBSStudio will allow the instructor (the developer) to overlay their image over a background showing a computer desktop or window.

Several quizzes will be created using the Moodle quiz tool to test the students knowledge along the way.

Appendix A contains a detailed list of the modules in each section that need to be produced.

Fallback Plan
Potential roadblocks have been identified that might divert this project in a different direction.

1) The combination of software versions may not be available or compatible with AWS. If this is the case, an earlier version of Moodle (3.0) will be used that works with earlier versions of the Linux-Apache-Mysql-PHP (LAMP) stack.
2) The Moodle installation may require more computing resources than an AWS Free Tier provides. If this happens, the author will pay the minimal amount needed to get sufficient resources from a more powerful AWS instance
3) Ubuntu 18.04 may not yet be available on AWS, in which case Ubuntu 16.04 will be used
4) Problems could arise with the computer and camera equipment used to record videos for MOOC content. The author has backup cameras and computers to use as needed.
5) The author currently has a free 5Mbps Google Fiber connection, which might be a bottleneck for uploading video content. The connection can be switched to a paid 100Mbps for short periods, then switched back to free, limiting costs.

LEARNING THEORIES BEHIND METAMOOC
This project involves the theory of technology acceptance as the project is rolled out in phases. Communication regarding the project is key to establishing positive and realistic expectations.

Situated learning theory is used as MOOC enrollees can communicate with each other and an instructor in social-media style forums as part of the Moodle installation.

Project-based learning is used as enrollees follow the examples given in sections teaching how to setup their own Moodle instance and how to navigate the maze of publicly available data sets.

MILESTONE 1
Milestone 1 will be comprised of the first two phases of the project: setting up the AWS instance, installing Moodle and documenting the process. Either a video trailer or the functional prototype of the Moodle lesson will be available at a publicly accessible website. A link will be provided to access this for peer feedback and mentor review.

Project plan may be adjusted as needed per feedback.

MILESTONE 2
Milestone 2 will cover phase three, which is creating the majority of the content related to finding and navigating publicly available data sets for MOOC researchers to use. This will essentially be a final project draft. A link will be provided to access this for peer feedback and mentor review.

Assuming these milestones go as planned, the first three phases will be complete at Milestone 2. Refinement of existing content from Phases 2 and 3, and/or creation of additional content identified as Phase 4 will go forward.

Project plan may be adjusted as needed per feedback.

APPENDIX A – LIST OF CONTENT / TASKS

PART 1 – Building a MOOC with Moodle Self-hosted (on AWS)
what you can do with your Free Tier AWS account
How to avoid being billed unexpectedly
spinning up an instance – EC2, Elastic IP, S3 storage (keep it free!)
connecting with SSH, AWS CLI and Linux basics
security groups, subnets, endpoints
other relevant AWS console features
Which versions of Ubuntu/PHP/mysql/Apache2 work with Moodle 3.1?
installing components required by Moodle
install/config moodle section
save your instance as a "golden master"

PART 2 – Populating Your MOOC
create content
upload and organize content
SEO considerations (important! To capture people searching for a Moodle MOOC offered in June by LearnMoodle
https://learn.moodle.net/)
considerations for registering students (full study beyond scope)

considerations for managing the course (full study beyond scope)

PART 3 – Sharing MOOC Data Sets
what data does Moodle generate?
What edX data is available?
ethical issues section
what other MOOC data is publicly accessible?
How to share your data

PART 4 – If all above is completed in <100 hours, do this…
expand content:
discuss windows and Mac install options
expand on Linux basics
link to other useful MOOCs
debugging Moodle self-hosted
MoodleCloud
edX platforms
udemy platform