

Corporate learning at scale: Lessons from a large online course at Google

Arthur Asuncion, Jac de Haan, Mehryar Mohri, Kayur Patel, Afshin Rostamizadeh, Umar Syed,

Lauren Wong

Google

76 9th Ave., New York, NY 10011

{arta, jacis, mohri, kayur, rostami, usyed, laurenbw}@google.com

ABSTRACT

Google Research recently tested a massive online class model for an internal engineering education program, with machine learning as the topic, that blended theoretical concepts and Google-specific software tool tutorials. The goal of this training was to foster engineering capacity to leverage machine learning tools in future products. The course was delivered both synchronously and asynchronously and students had the choice between studying independently or participating with a group. Since all students are company employees, unlike most publicly offered MOOCs we can continue to measure the students' behavioral change long after the course is complete. This paper describes the course, outlines the available data set and presents directions for analysis.

Author Keywords

MOOCs; connectivist MOOCs; corporate training; distance learning; online learning.

ACM Classification Keywords

K.3.1. Computer Uses in Education: Distance learning

INTRODUCTION

Massive Open Online Courses (MOOCs) have generated a great deal of excitement for their potential to make traditional university material accessible to a very wide audience. However, despite the growing popularity of MOOCs, there is considerable skepticism about their success and efficacy. A recent study showed that MOOC completion rates are very low, often in the single digits [2, 5], and some MOOC providers have shifted their focus to offering job training for corporations [1]. While there is an increasing focus on examining MOOC effectiveness in various contexts (see [3] for a recent example, and [4] for an excellent survey), relatively little analysis has been published in the corporate context. In light

of these developments, it is important to determine what features of a MOOC, if any, lead to the greatest educational gains in a corporate setting.

In the fourth quarter of 2013, Google Research produced a massive online class for company engineers around the world on the topic of machine learning with an emphasis on Google-specific machine learning software tools. We collected student-reported data about the course content and delivery method. As code written by Google engineers is stored in a central repository and all code execution is logged, we are able to directly measure the course's effect on student usage of the technologies taught. This rich data set may answer several important questions about the effectiveness of MOOCs: Are students who attended live lectures more likely to apply what they learned in the course than students who watched recorded lectures? Do a student's intentions to apply course content correlate with future actions? Does engagement with ancillary instructional channels (such as forums and office hours) have any impact on students' post-MOOC behavior?

COURSE DESCRIPTION

Approximately 6,500 students registered for an optional course on machine learning, representing a large fraction of all full-time employees distributed across more than 80 offices worldwide.

The 10-week course was a hybrid of theory-based lectures and Google-specific implementations. Each class was devoted to a single machine learning topic and was divided into three components. First, an internal machine learning expert delivered a lecture on the theory behind the weekly topic. The lecture was followed by one or more case studies, where experts explained how the techniques taught in lecture had been applied to solve important problems at Google. Finally, time was spent answering student questions from around the world. At the conclusion of each week, students were directed to an optional programming assignment to gain hands-on experience using internal libraries and technologies to reinforce core concepts from the class.

The course was offered in three formats. Each week for 10 weeks, a live video feed of the class was streamed to viewing rooms in many Google offices where students watched the content together in small groups and, in some cases, held dis-

cussions immediately after class. Students could alternately watch the live stream individually from their own computers. Finally, recordings of all 10 classes (as well as links to all slides, exercises, supplementary material and relevant external resources) were posted on the course website for asynchronous access.

DATA AND PRELIMINARY RESULTS

Per employee consent, all student page impressions were captured with a timestamp, length of page visit and unique student identifier. Individual student data was also captured when assignments were opened and code executed.

Most software written at Google is stored in a central repository, therefore this corpus can be crawled to locate files that reference machine learning function calls and algorithm libraries.

Students were surveyed 3 times over the 10 weeks. A pre-course survey was used to understand the prior level of experience with machine learning and students participation goals. A mid-class survey was used to gather feedback on class format and content pacing. A post-class survey was sent to collect student feedback for course improvements and understand how engineers plan to implement machine learning in future projects.

Forty-six percent of post-class survey respondents are “planning to use machine learning as a result of this class,” and six percent report that they are already “using machine learning as a result of this class.” Of final survey respondents, 62% report having machine learning conversations with others (managers, teammates, other students) as a result of the class.

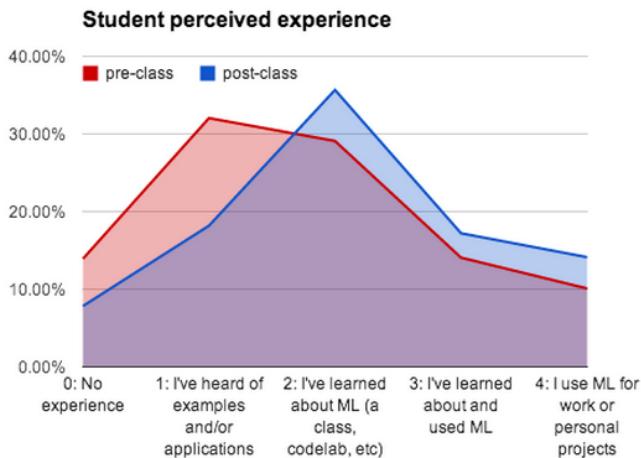


Figure 1. Comparing pre- and post-class surveys. Students responded to the question, “What is your current level of experience with machine learning?”

Future Research and Conclusion

Course designers will continue to collect and analyze longitudinal data over the coming year to assess course impact. Codebase references to machine learning files, individual satisfaction ratings and course participation data can be combined with employee-specific attributes (tenure, product area,

job title, level of education, etc) to answer our research questions:

- Which predictors (variables) are most likely to result in a student’s transition from learning to implementation?
 - What type of employees are more likely to adopt machine learning after taking the class?
 - Is there a difference in post-course satisfaction ratings based on a student’s reported pre-class experience with machine learning?
- Which course components were most successful?
 - Are students who attempted the interactive assignments more likely to implement machine learning?
 - As students begin using machine learning in real projects, does their perceived value of course components change over time?
 - Does the student-selected viewing method (independent or group setting, synchronous or asynchronous) have a measurable impact on course completion or future implementation of machine learning?
- Viewing Google as a social network, what is the density, reachability, and centrality of this content throughout the organization?

While the course served as a catalyst to build machine learning awareness, to raise visibility of leading experts and their work, and to foster dialogue across the company, the impact to engineering performance is yet to be determined. Rather than assessment of employee knowledge recall or recognition through online assessments, the *use* of machine learning concepts and tools in current and future products will become the measure of this course’s success.

REFERENCES

1. Chafkin, M. Udacity’s Sebastian Thrun, Godfather of Free Online Education, Changes Course. *Fast Company* (December 2013).
2. Jordan, K. MOOC Completion Rates: The Data. <http://www.katyjordan.com/MOOCproject.html>. Accessed: January 2, 2014.
3. Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of Third Conference on Learning Analytics and Knowledge* (2013).
4. Liyanagunawardena, T., Adams, A., and Williams, S. Moocs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distance Learning* 14, 3 (2013).
5. Perna, L., Ruby, A., Boruch, R., Wang, N., Scull, J., Evans, C., and Ahmad, S. The Life Cycle of a Million MOOC Users. Presentation at the MOOC Research Initiative Conference, December 5, 2013.