Massive Data, Individual Learners

Challenges for Developing Holistic Views of MOOC Participants

Paige Cunningham <pdcunni2@illinois.edu>
School of Information Sciences, University of Illinois at Urbana-Champaign

School of Information Sciences
The iSchool at Illinois

Introduction

What is a MOOC?

A MOOC is a Massive Open Online Course. MOOC courses began as a form of free online education, designed around individual stand-alone courses, but have evolved to include sequential, paid, for-credit programs as well.

Why are MOOCs significant?

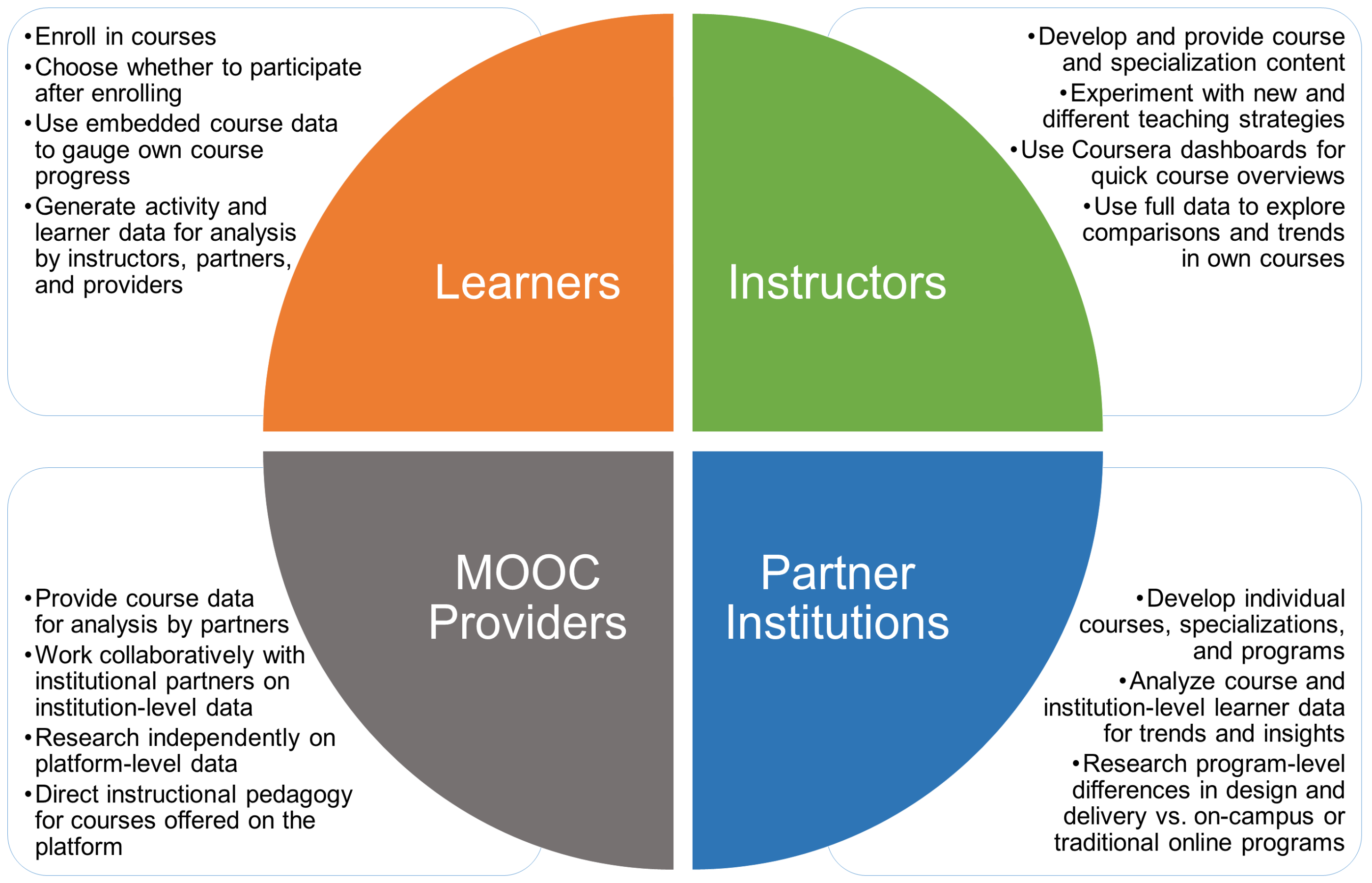
The enrollment numbers for MOOCs can be immense, which means that enormous amounts of learning-related data is being generated and made available to the University of Illinois and Coursera, our MOOC partner.

What are the Challenges?

Not all of the generated learning-related data is equally meaningful or useful, not all potentially useful data is made available to the University of Illinois, and the data that is available has changed significantly over the four years that the University of Illinois has been partnered with Coursera, making longitudinal comparisons of course offerings and learner patterns across the delivery modes very difficult.

Stakeholders

As MOOCs are partnerships between institutions and providers, they have additional stakeholders beyond traditional campus-based courses, each with their own concerns and data needs.



Learners
- Enroll in courses
- Choose whether to participate after enrolling
- Use embedded course data to gauge own course progress
- Generate activity and learner data for analysis by instructors, partners, and providers

Instructors
- Develop and provide course and specialization content
- Experiment with new and different teaching strategies
- Use Coursera dashboards for quick course overviews
- Use full data to explore comparisons and trends in own courses

MOOC Providers
- Provide course data for analysis by partners
- Work collaboratively with institutional partners on institution-level data
- Research independently on platform-level data
- Direct instructional pedagogy for courses offered on the platform

Partner Institutions
- Develop individual courses, specializations, and programs
- Analyze course and institution-level learner data for trends and insights
- Research program-level differences in design and delivery vs. on-campus or traditional online programs

Acknowledgments

The author thanks Dr. Maryalice Wu and the CITL Data Analytics Team for all of their work on the University of Illinois Coursera data and Dr. Linda Smith from the School of Information Sciences for her support and encouragement.

ILLINOIS

MOOCs @ Illinois

Between August 27, 2012 and August 26, 2016, the University of Illinois offered 66 **Unique Courses** through Coursera. These courses are spread across two different platforms, the older Session-Based platform having been fully replaced by the newer On-Demand platform in 2016.

Course Type	Session-Based	On-Demand
Dates for Course Offerings	**08/27/2012 – 02/28/2016** (Platform closed)	**04/21/2015 – *08/26/2016*** (Courses are on-going)
Unique Courses	**24** (15 moved to On-Demand)	**57** (15 moved from Session-Based)
Specializations	**2** (2 moved to On-Demand)	**11** (2 moved from Session-Based)
Enrollees	**1,389,748**	**492,285**
Active Participants	**701,856** (50.5% of Enrollees)	**286,806** (58.3% of Enrollees)
Certificate Earners	**45,355** (6.5% of Active Participants)	**25,388** (8.9% of Active Participants)

Many of the newer on-demand courses are being deployed as **Specializations**, sequences of courses which qualify for specialization certificates. Some specializations also meet the requirements of specific academic degree programs.

Enrollee counts are based on unique enrollments in unique courses, that is learners enrolled in more than one session of a Session-Based course are only counted once. As many courses have moved between platforms, learners may be included in the enrollee counts of both platforms.

Active Participants are enrollees who have watched at least one lecture, taken at least one quiz, or contributed to at least one forum.

Certificate Earners are active participants who have qualified for a course completion certificate by meeting the course requirements.

Data Types

- **Lecture Videos**
  - **Session-Based** course videos were open to all learners. Data included information on both streaming videos actions (including pauses, skips, etc.) and video download data (download counts only). No data was provided on mobile app lecture interactions.
  - **On-Demand** course videos are open to all learners. Data is far more limited, with streaming video data only reporting completion and view counts, and no download or app data available.
- **Forum Participation**
  - **Session-Based** forums were open to all learners. Data included information on who started threads, who responded to threads, and post content.
  - **On-Demand** forums are open to all learners. Data includes information on who starts threads, who responds to threads, and post content.
- **Graded Assessments**
  - **Session-Based** graded assessments were open to all learners. Data included information on both graded assignments and summative quizzes and exams for all active participants.
  - **On-Demand** graded assessments are either open to all learners or restricted to Paid Learners. If Premium Grading in effect, non-paid learners are considered “Auditors” and do not have access to graded assignments or summative quizzes or exams, so no assessment data is available about them.
- **Clickstream Data / Activity Logs**
  - **Session-Based** courses provided clickstream and activity log data for all learners. Data included information on what pages learners looked at, learner IP address information, and streaming video interaction data.
  - **On-Demand** courses currently do not provide clickstream or activity log data. Clickstream data is reportedly forthcoming, but is currently missing.
- **Survey Data**
  - **Session-Based** surveys asking about demographics, intention, goals, and satisfaction were distributed by our group at the beginning and end of courses.
  - **On-Demand** surveys about demographics, intention, goals, and satisfaction are currently distributed by Coursera based on triggers such as three weeks of non-participation, as there is no clear “end” to on-demand courses.

Data Challenges

Learner grade data is relatively easy to extract, but provides only a superficial understanding of learner engagement. More meaningful glimpses into such engagement can be acquired from in-depth lecture watching and clickstream data, when it is available, as well as from forum and survey data, though forum data is typically messy and hard to analyze in a timely fashion and survey data tends to be very sparse.

In addition, the transition between analyzing Session-Based and On-Demand course data has been rocky. Both the Session-Based and On-Demand modes ran simultaneously for nearly a year, emphasizing the difficulties caused by the shift.

We typically received Session-Based data one month after each Session-Based course closed. In contrast, we only acquired access to On-Demand data in late 2015, more than 6 months after our first On-Demand courses started. Once we began receiving On-Demand data we learned that some data sources are still not available for On-Demand courses, and other data is presented in very different ways.

Finally, many on-demand courses are now using the Premium Grading option, which enforces an Auditor/Paid Learner split. In these courses learners can only access the required assessment materials if they pay for a certificate, impacting participation rates.

Free certificates were available in all session-based courses (with a paid Verified Certificate option added in early 2014), but most on-demand classes now use Premium Grading, in which only participants who have paid for the course are able to see the materials that are required for course completion, eliminating the option for free certificates.

Identifying course “drop outs” is also particularly difficult, as only Paid Learners can officially complete the course, leaving both non-completed Paid Learners and Auditors who have unofficially left the course floating as “phantom course members” due to an inability to mark themselves as finished without reaching the completion milestone.

Implications

The University of Illinois is spending significant time and effort on developing MOOC courses, and has even created graduate degree programs based around Coursera. The College of Business’ iMBA program officially enrolled its inaugural cohort in January of 2016. A second program, the Department of Computer Science’s Master of Computer Science in Data Science (MCS-DS), launched in August of 2016.

Knowing what learners do and don’t do, who they are, why they may have stopped participating in a course, etc. can help make sure that all learners, paid and unpaid, get the best possible education and experience in all of our Coursera courses.

Unfortunately, it can be difficult to make these kinds of judgements based solely on data available directly from Coursera, especially when the available data changes significantly, as happened with the switch from the Session-Based to On-Demand platforms. Surveys are being used to help flesh out the available data, but are still limited in what they have been able to establish. Integrating different data sources makes defining an overall picture of learner engagement more difficult, though the process can provide granular pictures of individual students’ participation in individual courses.

Often the people who handle the data are not the people who can make the assessments about the outcomes, and so we need to progress toward working in tandem between the two groups. All stakeholders need to work together to help the right data get to the right people in order to ensure individual learners get the best possible experience from the courses offered on the platform.

For more information on the University of Illinois’ Coursera data, visit: **mooc.illinois.edu**