Learning Dashboard: Bringing Student Background and Performance Online

Fatemeh Salehian Kia
Simon Fraser University
fsalehia@sfu.ca

Zachary A. Pardos
UC Berkeley
zp@berkeley.edu

Marek Hatala
Simon Fraser University
mhatala@sfu.ca

Abstract

Massive open online courses (MOOCs) generate large quantities of rich data, providing abundant opportunities for educational research. Data attributes of student background and courseware interaction logs provide information that can be predictive of a struggling or succeeding student but these signals are often lost in the data heap and never get surfaced to instructors. In this paper we build on prior work and showcase an instructor analytics module that puts the user in charge of data inquiry. Our module is open-source and compatible with moocRP, an open source open learning analytics platform. Using data from an edX course, we investigate learner performance correspondence with attributes such as; age, gender, and country of origin. This paper proposes an interactive visualization filtering, which enables studying likely impacts of one indicator to the other. Demographics, most notably age and country, point to potential differences in students' motivations for participating MOOCs. Thus, this tool can inform educators and course moderators about these differences for further design improvements.

1 Introduction

Digital learning environments, and in particular, massive open online courses (MOOCs), strive to retool pedagogy to improve self-regulated learning by increasing and tracking students' engagement. Current generation MOOC platforms such as edX, Coursera, and Udacity, which offer free online courses, draw students with diverse backgrounds from nearly every country in the world. While instructors can learn about their students' educational background and experience as they interact with them in the traditional classrooms, MOOC instructors do not have this opportunity. Consequently, this sort of online learning path and content that current MOOCs offer potentially make them ill-suited for non-autodidactic learners [Mcl13]. However, MOOCs are able to collect information about students' interaction with the platform as well as background information by asking students to answer a survey when they sign up for the course. This information when used formatively can help adapt the courseware for students in need.

There are a few conventional research studies conducted on data analysis of students' demographics related to their achievement [Bre13, Deb13, Guo13, Duv11, Xu14]. Breslow et al. (2013) investigated the survey and clickstream logs of the first edX MOOC to produce an in-depth picture of enrolled students. Though they

In: A. Wolff, Z. Zdrahal (eds.): Proceedings of the Data Literacy for Learning Analytics Workshop, Edinburgh, UK, 26-Apr-2016, published at http://ceur-ws.org

provided visualizations to demonstrate indicators such as age distribution, reasons for enrolling, etc., these visualizations are static and built for specific research questions which may not align with questions instructors may have.

In current paper, we address the question of how students' attributes such as age, gender, location, and educational background relate to their performance and achievement. These data generated by MOOCs can be represented on the dashboard to help instructors consider instructional strategies and interventions. For example, which ways of content delivery support or do not support understanding of a complex concept [Bre13]. Thus, to provide a clearer picture of students to help educators and course moderators learn more about their target learners, we augmented and developed an interactive visualization dashboard for moocRP [1] [Par16] to aggregate and visualize data. moocRP is an open-source, web-based data repository and analysis tool tailored to the MOOC community. As a proof of concept, we used edX clickstream and survey logs of the introductory Statistics course, which began in the Spring 2013 to show how the visualization dashboard works. With moocRP, institutions needing control over data distribution do not need to upload data to a centralized location; instead they can import and run the desired analytic module locally [Par16].

2 Visualization Indicators

Since MOOCs are relatively new, there are few research studies that have been conducted on visualizing rich information available in the dashboards. Martinez et al. (2014) administered a survey to investigate which information sources are most useful for MOOC instructors. The result confirmed earlier surveys that overall student performance relative to the class is one of the most important information for the instructors [Mar14]. Although visualizing information of every single learner performance in MOOC is not feasible, learners' attributes such as demographics or educational background enable instructors to analyze performance of certain group of learners, for instance, performance of learners who are in their 20s. This is in line with educational research indicates these latter variables are important controls in predictions of educational outcome [Bre13]. Thus, performance as well as demographics and educational background indicators were depicted in the interactive dashboard. All charts and histograms are rendered using *D3* [Bos11]. *Crossfilter* is a *JavaScript* library, which was used to enable multivariate filtering [Cro16]. The filtering feature makes this dashboard distinct from similar interactive dashboards.

2.1 Demographics and Locations

Our edX MOOC dataset includes student answers to a survey developed to learn more about student backgrounds who signed up for an account. The survey questions included not only demographics such as age and gender, but asked students, for example, about their educational and professional background. Based on the survey answer, which we call profile logs, student's demographics were rendered by histograms.

- *Age*: is calculated by the year of birth. Ages that are less than 10 and greater than 80 are filtered out.
- *Gender*: is rendered according to values in the data set, since they might be defined differently in various MOOCs profiles.
- *Level of education*: is also depicted by capturing the values from the data set.
- *World map of enrolment*: shows the locations from which students accessed the course site. The student's IP address was recorded each time he or she interacted with the website. The choropleth map, where different countries color-coded, was built to demonstrate the distribution of enrolled students in the course.

2.2 Performance and Achievement

Building from the work described above, the second phase of study sought to visualize usage data and achievement (e.g. certificate earners) in the dashboard. This information could potentially identify relationships between the characteristics and capabilities of the students themselves and their achievement. As mentioned above, the most important information for instructors are the overall student performance [Mar14]. In MOOCs, courses

[1] https://github.com/CAHLR/moocRP

normally include weekly modules. Each module consists of a set of videos. Interspersed among the videos are online exercises that give students an opportunity to practice the concepts covered in the lecture videos. These exercises as well as assignments are graded. Students need to accrue a minimum of points to earn a certificate. Based on this course structure, three indicators were developed in the dashboard to demonstrate students' performance and achievement:

- *Attendance in modules*: shows a number of learners navigating through the course module.
- *Percent of correct*: calculates the overall percent of correct answer to the exercises within each module.
- *Certificate earned*: demonstrates a number of students who received or did not receive a certificate of completion.

As a proof of concept, we used data generated by a UC Berkeley MOOC, "Introduction to Statistics" on the edX platform. These data include the IP addresses of all enrolled students; clickstream data that recorded each of the 43 thousands interactions the students had with the platform; and the profile information of enrolled students. To preprocess the data sets, login location countries were determined by using geolocation database [Geo16]. If a student logged in to his or her account with the IP addresses from different countries, the most frequent country was chosen. Finally, the redundant data were removed and all three data sets were aggregated to a single one.

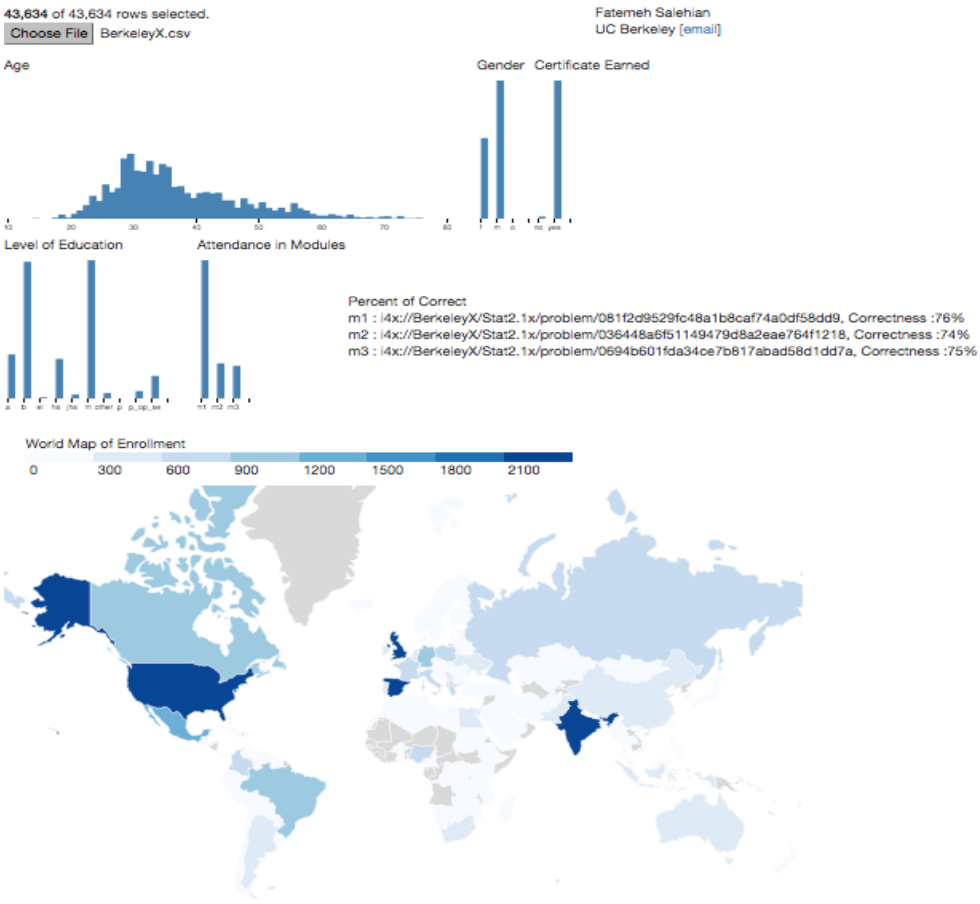

Figure 1: The Learning Dashboard

Figure 1 depicts a screen shot of the learning dashboard using the aggregated data set. The dashboard is interactive. Hovering over every country pops up the tooltip with the country name and number of enrolled students inside. In our dataset, students came from 194 countries throughout the world. The top three countries were the United States (9275), India (5434), and the United Kingdom (2265). The highest populated country, China, was responsible for only 404 enrolled students in our dataset.

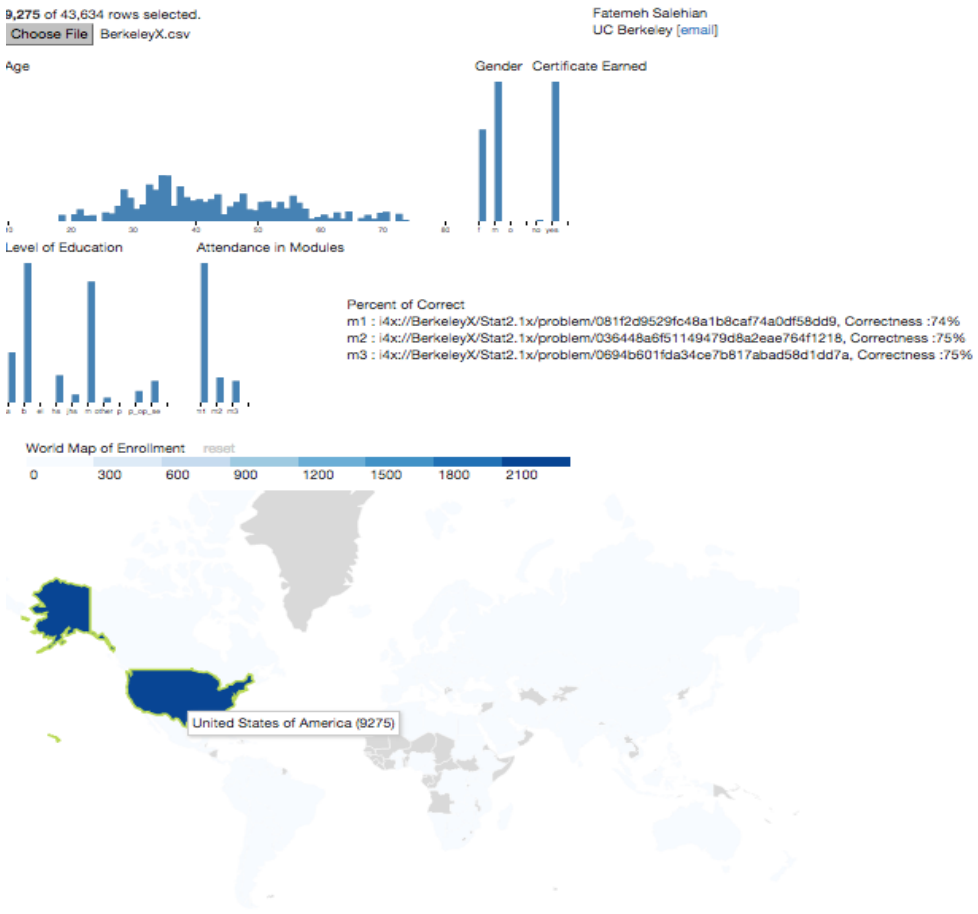

Figure 2: The Learning Dashboard, Map Filtering

The filtering feature of the dashboard makes it distinct from similar dashboard offerings. Figure 2 demonstrates how choropleth map filtering works by clicking on a specific country. The *jQuery* plug-in called `datamaps` and *Crossfilter* library were used to develop this filtering feature. Similarly, sliding the brush over the columns can filter the charts. The brush size is flexible and filters the chosen space, so that a user can customize the width of filtering. Furthermore, multivariate filtering is also possible in the learning dashboard. That is to say, multiple visualization indicators can be filtered simultaneously. Figure 3 shows this feature of the dashboard.

3 Initial Evaluation

Several dashboard applications have been recently developed to support learning or teaching. These dashboards are used in traditional face-to-face teaching, online learning, or blended learning settings. Most of these dashboards are deployed to support teachers to gain a better overview of course activity. But few of them evaluate the actual impact on inducing new understanding [Ver13] Said another way, the features and usability of dashboards often are misaligned with expectation of the user.

We conducted a preliminary qualitative user study with 5 participants in an attempt to examine the usability of the dashboard. Two of five participants taught MOOCs, and three of them are instructors in blended learning courses, or they have experience working with MOOC data. After a one-minute demonstration of how the dashboard works, they were asked to interact with the dashboard for a minute. Then study participants completed a questionnaire that asked about what kind of information they found interesting to learn from interacting with the dashboard, and what kind of information was missing. Table 1 is a list of "use" and "requests" gathered in the initial user study. Most of the participants found the dashboard intuitive and easy to use. The multivariate filtering is a feature that they found useful for analyzing one indicator over the others. Four out of five participants found demographics information useful for improving the course content. One of the participants

found the comparison of performance of learners by their location informative. For instance, if the percent of correct answer of a course module for a certain country is lower than the average percent, it might show the lack of prerequisite in their national curriculum for the subject. However, all of participants requested more specific information about learners' interactions and performance with the courseware.

Participants were also asked about what kind of information they get from in-person classes that they cannot get or it is hard to get from online courses. The majority of participants responded that in general they are not aware of students motivation in completing online courses. Moreover, instructors are not fully informed about how students perceptions about features of content and of the instruction influence their performance in an online course.

4 Discussion and Conclusions

In conclusion, this paper has presented a MOOC dashboard for considering learners' attributes in online courses, with a primary use case being to examine what factors of students' demographics might be correlated with their performance and success. These analytics could inform educators and course moderators, and help complete the Learning Analytics Cycle. According to Clow [Clo12], this cycle conceptualizes successful learning analytics work as four linked components: (1) learners (2) generating data (3) that is used to produce metrics, analytics or visualizations. The key step is 'closing the loop' by feeding back this product to learners or people involved in the learning process through one or more interventions (4). Based on Laurillard's Conversational theory [Lau13], learning takes place through a series of 'conversations' between a teacher and a student (and with other students). Thus Learning Analytics Cycle can be conceptualized as enabling conversations by translating information from students to instructors via analytics in the form of interventions [Clo12]. In the context of MOOCs, we strove to carry on these conversations between learners and instructors in online settings by providing visualizations to instructors to support their interventions. Thus, instructors and course moderators are informed about potential differences in students' motivations for participating MOOCs for further course design improvements.

In this paper, we developed and augmented the visualization dashboard for moocRP. The application of filters facilitates exploring the large multivariate data sets in an intuitive fashion with fluid presentation. However, this dashboard needs to be augmented by more sophisticated predictive analysis, as well as, more specific visualizations about learners' performance and activities informed by further user study in the future work.

References

[Par16] Pardos, Zachary A., Anthony Whyte, and Kevin Kao.moocRP: Enabling Open Learning Analytics with an Open Source Platform for Data Distribution, Analysis, and Visualization. *Technology, Knowledge and Learning: 1-24*

[Bre13] Breslow, Lori, et al. Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research Practice in Assessment 8*, 2013.

[Deb13] DeBoer, Jennifer, et al. Bringing student backgrounds online: MOOC user demographics, site usage, and online learning. *engineer 2: 0-81*, 2013.

[Mcl13] McLoughlin, Catherine Elizabeth The pedagogy of personalised learning: exemplars, MOOCS and related learning theories. *Proc. EdMedia*, 2013.

[Guo13] Guo, Philip J., and Katharina Reinecke Demographic differences in how students navigate through MOOCs. *Proceedings of the first ACM conference on Learning@ scale conference. ACM*, 2014.

[Mar14] Stephens-Martinez, Kristin, Marti A. Hearst, and Armando Fox. Monitoring MOOCs: which information sources do instructors value? *Proceedings of the first ACM conference on Learning@ scale conference. ACM*, 2014.

[Duv11] Duval, Erik.Attention please!: learning analytics for visualization and recommendation. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge. ACM*, 2011.

[Ver14] Verbert, Katrien, et al.Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing 18.6: 1499-1514*, 2014.

[Xu14] Xu, Zhengzheng, et al.Visual analytics of MOOCs at maryland. *Proceedings of the first ACM conference on Learning@ scale conference. ACM*, 2014.

[Geo16] GeoIP Databases Services: Industry Leading IP Intelligence. (n.d.). Retrieved from `https://www.maxmind.com/en/geoip2-services-and-databases`

[Bos11] Bostock, M., Ogievetsky, V., Heer, J. D3 data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on, 17(12), 2301-2309*

[Cro16] Crossfilter: Fast Multidimensional Filtering for Coordinated Views Retrieved from `http://square.github.io/crossfilter/`

[Clo12] Clow, Doug. The learning analytics cycle: closing the loop effectively. *Proceedings of the 2nd international conference on learning analytics and knowledge. ACM, 2012.*

[Lau13] Laurillard, Diana. Rethinking university teaching: A conversational framework for the effective use of learning technologies. *Routledge, 2013.*

[Ver13] Verbert, Katrien, et al. Learning analytics dashboard applications. *American Behavioral Scientist (2013): 0002764213479363.*

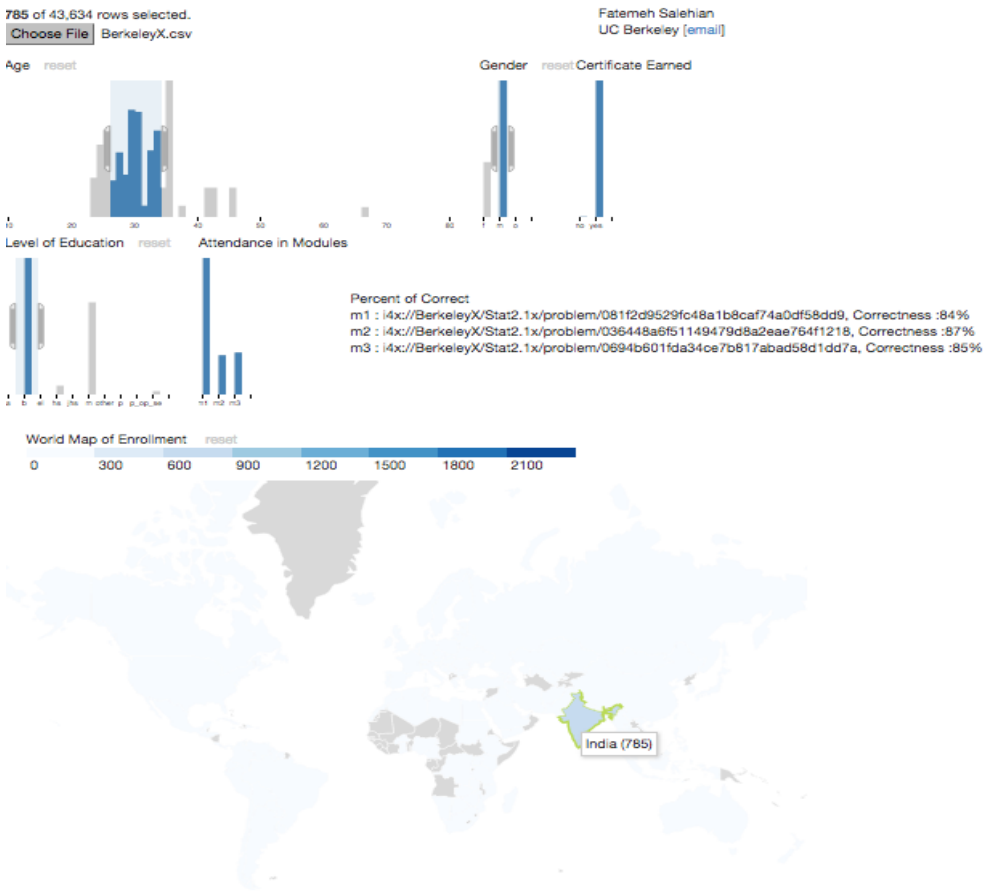

Figure 3: The Learning Dashboard, Multivariate Filtering

Use	Request
Opportunity to explore patterns of demographics that relate to attendance or success.	1. What information do students operate on e.g. highlight or refer to in a note they make? 2. What specific information do students' review? What are temporal features of review e.g. cramming all at once or reviewing a bit at regular intervals?
The multivariate application of filters was good. The demographic differences over learners' performance were informative.	I would like to see more information about learner performance and learning activity, for example trend of students navigation through the course components.
The application of filters, multivariate filtering, intuitive.	More specific information about "percent of correct" e.g. for each question, percent of dropout
How you can narrow down the information for specific category of learners. The dashboard is easy to use.	More specific information about learners' interactions with the website. Example: number of attempts for answering questions, viewing videos (top videos), etc.
This was not very informative, as I do not find gender and location particularly helpful.	Individual differences over learning outcome

Table 1: List of "uses" and "requests" of learning dashboard (Pros and Cons)