

The University of Southampton MOOC Observatory Dashboard

Manuel León-Urrutia¹ and Darron Tang¹

¹ University of Southampton. University Road SO172BJ, UK

{m.leon-urru^tia,darron.tang}@soton.ac.uk

ABSTRACT. The University of Southampton MOOC Observatory Dashboard (UoSMOD) is an application that visualises near-to-real time data from FutureLearn courses. The intended end users of this tool are those who are involved in MOOC development and delivery such as mentors, educators, learning designers, researchers, programme leaders, and marketing officers. These different stakeholders (mentors, educators, learning designers, etc) are beneficiaries of different features of UoSMOD, who use them for different purposes. The tool downloads the data dumps that FutureLearn provides to their partners every 24 hours, and scrapes the courses metadata from the administration site of the platform. The data is managed in a MySQL database, and an R based environment called Shiny is used for its analysis and visualisation. These visualisations have been presented to mentors and learning designers. New features have been being added as a response to the feedback provided by its first users. Further iterations are in the pipeline, in this process of optimising a tool that exploits the data available in the most usable way as possible.

Keywords: MOOCs, Dashboards, Visualisation, Learning Analytics.

1 Introduction

Massive Open Online Courses (MOOCs) produce large quantities of data, the full potential of it is yet to be exploited [1]. Insights obtained from MOOC data can be used for purposes such as personalisation [2], performance prediction [3], and curriculum improvement [4].

However, obtaining meaningful insights from MOOC data is challenging because not all data analysts are fully aware of the design and intent of the courses to be analysed, and those who are acquainted with the context do not always have the time or skills to analyse such data.

In order to address such gap, MOOC platforms offer visualisation dashboards of part of their data, together with the data itself. For example, platforms such as EdX and

FutureLearn data: what we currently have, what we are learning and how it is demonstrating learning in MOOCs. Workshop at the 7th International Learning Analytics and Knowledge Conference. Simon Fraser University, Vancouver, Canada, 13-17 March 2017, p. 8-19.

Copyright © 2017 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. This volume is published and copyrighted by its editors.

Coursera provide their partner universities with visualisations of their courses data. These visualisations provided by MOOC platforms are aimed at making accessible the analysis and visualisation of MOOC data to educators. The visualisations are based on the data that they supply to the partners, but partners often demand more visualisations than those supplied. Also FutureLearn provides a summary table with figures for each of their courses, including number of comments, total enrolments, visited and completed learning activities, and number of learners according to different levels of engagement. FutureLearn also provides a facilitation dashboard, which identifies steps by number of comments, and sorts comments according to their impact in the course (number of responses, number of likes). In order to address such shortage, several attempts have been made at developing tools that provide visualisations beyond what the platforms can offer. For example, Cobos et. al. [5] developed Open DLAs, a plug-in that visualises data from EdX in more detail than what the platform offers. Also Chitsaz et. al. [6] developed a tool that provides further visualisations to those offered by FutureLearn, with similar aims to the present project, namely the University of Southampton MOOC Observatory Dashboard (UoSMOD): providing visualisations with finer grain than those provided by the MOOC platforms, based on the same datasets.

2 The University of Southampton MOOC Observatory Dashboard: The Data Analysed

When interacting with the platforms where MOOCs are hosted, learners leave a significant amount and variety of digital footprints that remain recorded in the platform database. The provision of some of this data for evaluation and research purposes is usually part of the agreement between universities and platforms. In the case of FutureLearn, the data is provided in a set of datasets in csv format. All datasets have two data types in common: a unique anonymised identifier for each learner, and a timestamp. The datasets are the following:

- *Enrolments*: Each user who signs up to the course is registered with a unique identifier, and the date of enrolment is recorded. If the user leaves the course, a leaving date is recorded too.
- *Demographic data*: This data is integrated within the *enrolments* dataset. The demographic data is the result of a survey that the platform runs at the beginning of each course.
- *Comments*: Each comment made by each user is recorded with a timestamp, a comment ID, and its author unique ID. If it is a reply, the ID of the parent comment is also registered. There is also an indication of how many likes the comment has received, and whether it has been moderated.
- *Step activity*: A record of each time a user visits a learning object for the first time, and each time a user marks it as complete.

- *Quiz results*: The results of the multiple choice questions attempted by each learner, in case there are multiple choice questions in the course.
- *Peer review exercises, and reviews*: All texts produced by students in the peer review activities, if they exist in the course.

3 The UoSMOD: Data Retrieval and Display

FutureLearn updates these datasets every 24 hours. The files are available for download in the admin page of each course, within the admin site of each partner, as shown in figure 1:

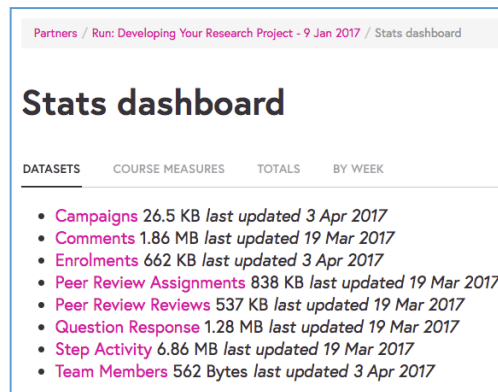


Fig. 1. List of updated datasets supplied by FutureLearn for a run of a MOOC.

As seen in the figure above, the reduced size of the datasets allows downloading a high quantity of them without great server demands.

This is achieved using a web scraping Python script using the BeautifulSoup library. The script downloads these files and obtains other metadata from the admin page, such as title, start date, and run number. Both data and metadata are combined and converted into SQL, and transferred to a MySQL database. A web application called Shiny (an R based environment) is used to analyse and visualise the data from the SQL database, and displays it in the dashboard. The process is represented in figure 2:

The implementation requirements of the application are the following:

1. An Ubuntu server with a minimum of 2GB of RAM
2. The configuration of R-Studio with up-to-date packages (a detailed list can be found in <https://github.com/moocobservatory/mooc-dasboard/>)
3. Installation of Shiny Server (the Open Source version is sufficient)
4. Installation of MySQL, with a root password
5. Configuration of Shiny Server
6. Configuration of the Shiny Dashboard

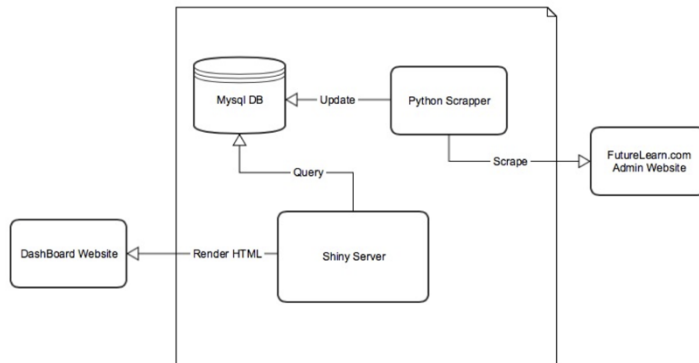


Fig. 2. The UoSMOD process

4 Features

The result of this process is a dashboard with several features, the choice of which was initially made by the learning designers of the MOOCs in Southampton, in response to their needs to understand the effectiveness of their learning design. These features are in constant evolution, as new features are being incorporated and other features are modified in response to a larger pool users' feedback, including educators and mentors. At present, the tool provides the visualisations described in the following sections.

4.1 Multiple course selection

The dashboard contains an interface to simultaneously select up to four different runs of different courses for comparison, which offers different metrics such as demographics, step activity, and comments, as figure 3 below shows

The screenshot shows a user interface for selecting courses. At the top, a message reads: "Loading data for all courses may take a few minutes." Below this, there are four "Course selection" sections. Each section contains a "Courses" dropdown menu and a "Run" dropdown menu. The first section has "Agriculture 1415: Myth and Reality" selected in the "Courses" dropdown and "1 - 2015-10-19 - 2015-11-02" selected in the "Run" dropdown. The other three sections have "None" selected in both dropdowns. To the right of the first section is a "Go" button. To the right of the other three sections is a "Go" input field.

Fig. 3. Course selection interface

4.2 Aggregate enrolment data

This feature is aimed at strategic stakeholders, as all measures of all courses are combined in the same table.. It aggregates relevant data from all runs of different courses, such as statements sold, enrolled learners, completers, leavers and social learners. The table that can be downloaded in a csv file for further analysis, and in pdf for a printout.

4.3 Demographics

The results of the surveys launched in each run are presented in different bar charts and maps. Figure 4 below shows two compared courses, selected with the course selection interface, one represented in blue, another represented in black. The metrics shown in the figure are age ranges, gender, and profession. The data of all metrics can be also downloaded in a csv file. A filter of the demographic metrics by learners who purchased a statement is also available, for market research.

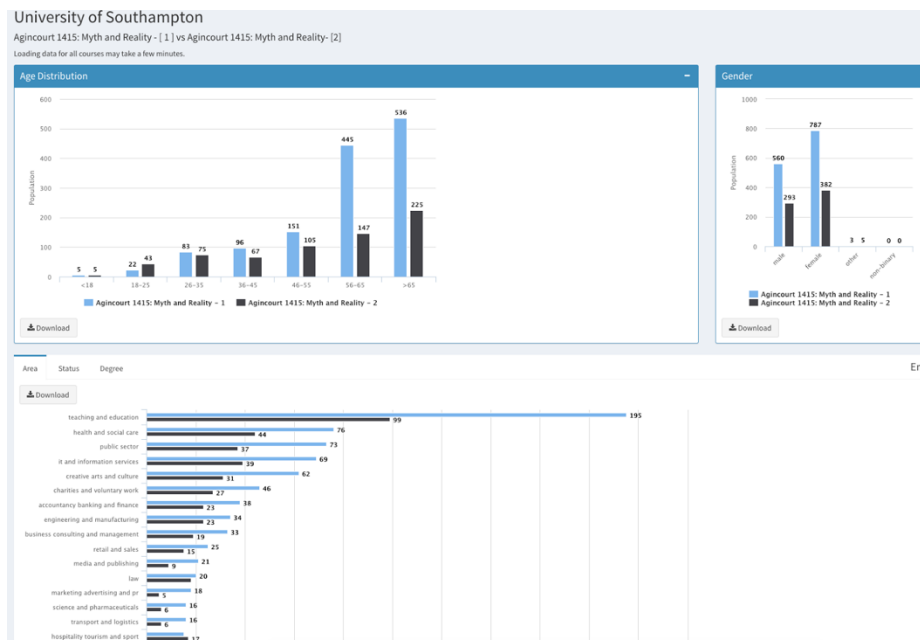


Fig. 4. Comparison of different demographic metrics

4.4 Registrations and statements sold

This visualisation shows sign ups and statements sold over time before, during and after the course runs. Figure 5 below shows two histograms. The histogram at the top shows the enrolments before and after the course. The line in the middle is the start date of the course. The histogram at the bottom shows the statements sold over time, since the first day of the course.

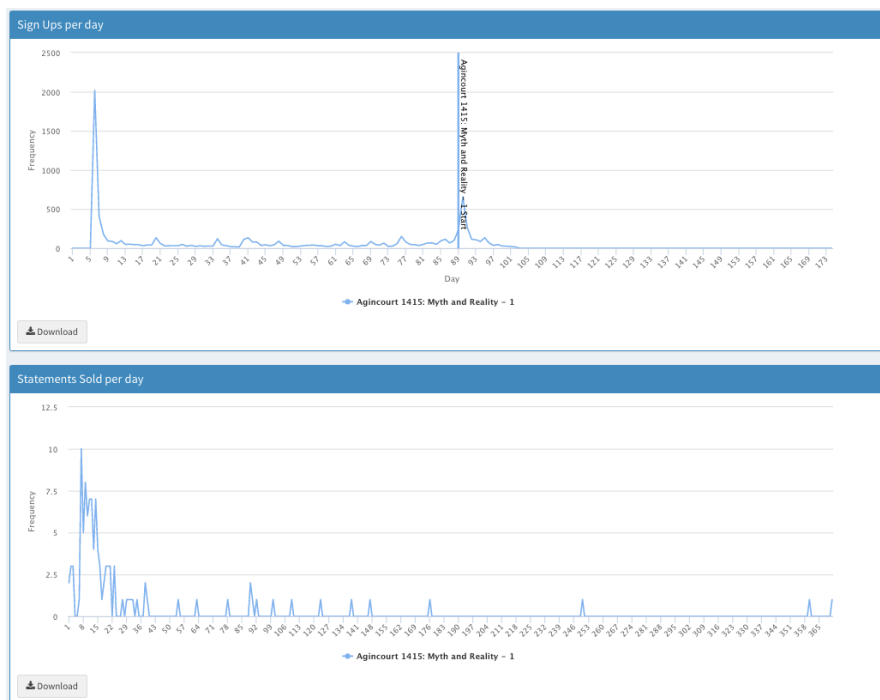


Fig. 5. Sign ups and statements sold histograms

4.5 Step completion

This section shows three visualisations (see figure 6 below). The bar chart at the top shows the comments made in each of the steps of the course. The heatmaps in the middle and the bottom show comments first visited and completed by step and date respectively. It is possible to hover the mouse over every cell, which will provide information about date, step, and number of events (see grey box in the middle of the top heatmap).

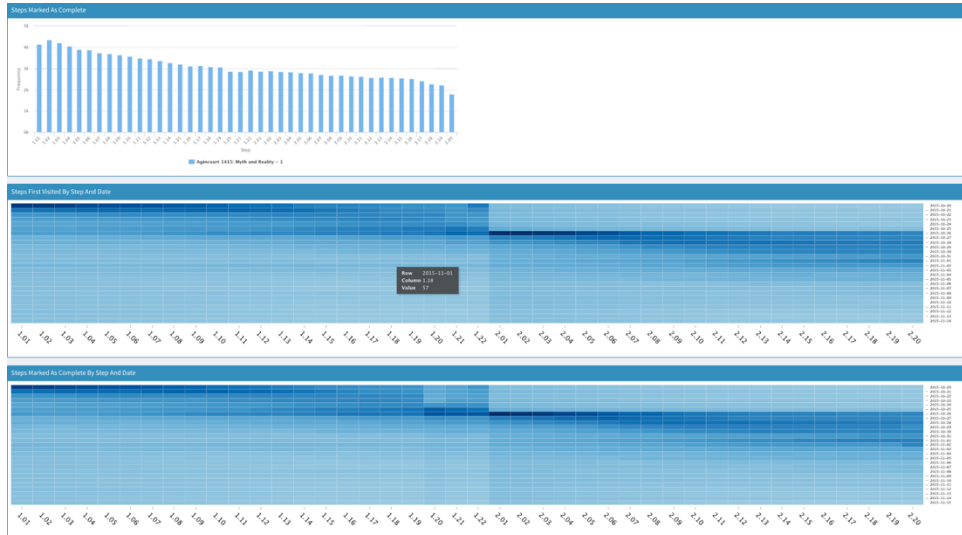


Fig. 6. Step completion metrics

4.6 Comments overview

This feature provides visualisations of the numbers of comments and responses made during the course. The bar chart at the top of figure 7 shows the number of comments (in black) and replies to comments (in blue) made in each of the steps of the course. The heatmap in the middle shows the number of comments per step per day. A particular step can be highlighted (see vertical highlighted line in heatmap), as well as a particular day (see horizontal line). Finally, the bar charts at the bottom of the figure show comments and replies per week on the left hand side, and authors per week on the right hand side.

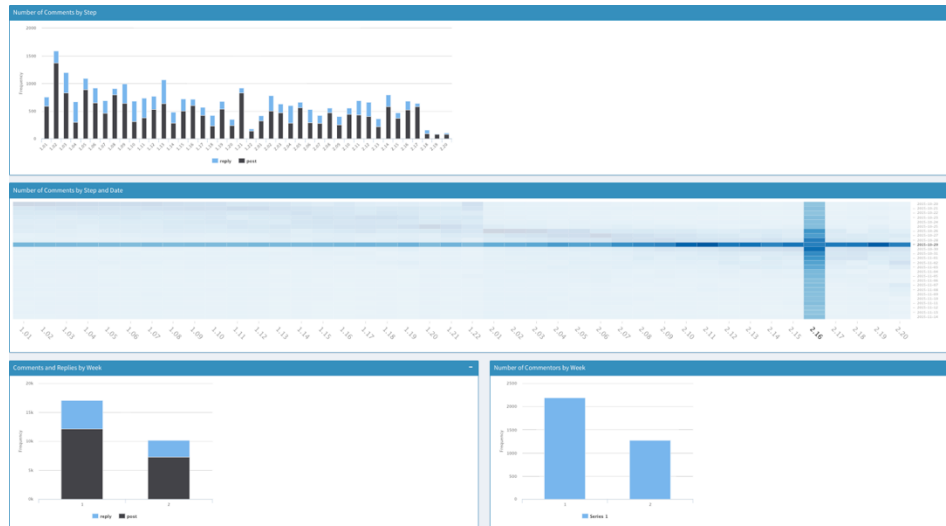


Fig. 7. Comments metrics: comments and responses (above), comments heatmap (middle), comments per week (below)

4.7 Comments viewer

This feature filters and sorts comments by dates, steps, keywords, and provides context to them, such as whether they belong to a thread, or how many likes they received. This tool is aimed at providing quick access to relevant comments so that they can be addressed accordingly. For this, there is also a link back the platform that brings the user straight to the comment in question (last column of table in figure 8). The tool also provides a word cloud (see top right box in figure 8) of the filtered comments, that can be adjusted in terms of frequency and number of words (slide bars at the top left of the figure)



Fig. 8. Comments viewer

4.8 Correlations

This feature allows seeking for correlations of different measures of a course. The course can be selected from the box at the top of Figure 9, and the measures to compare from the two boxes below the course selector. The measures that can be compared are:

- Number of comments
- Number of replies
- Number of likes
- Number of submitted quiz responses
- Percentage of correct quiz responses
- Percentage of incorrect quiz responses
- Number of completed steps

These are placed in an x and y axes, and a scatter plot with a regression line is returned. In the figure below, a positive correlation is shown between learners of a course number of comments and the number of completed steps in a part. Each dot represents a learner, and hovering the mouse on it will provide the exact measures for that student.



Fig. 9. Correlation of number of comments per percentage of step completed made by every learner

5 Outcomes

The tool is regularly used by certain stakeholders. For example, the programme lead is finding very useful the aggregate measures feature, for producing reports to the senior management and to other stakeholders who require a quick snapshot of the measures of particular courses. This feature is also proving very useful to keep track of the sales made by the different courses of the university programme, and other measures such as the registrations before the course starts so that support strategies can be planned accordingly..

The enrolments feature is also occasionally used by the programme leader, and several lead educators and learning designers for strategic purposes, especially to seek for

evidence of impact of campaigns and other events that may have an impact in the enrolment figures. This can help making informed decisions as to when is the best date for opening a course, or to quickly compare courses in terms of their enrolment figures.

Mentors use the comments viewer as a feature that allows mentors to quickly find questions by filtering by question marks and other keywords. They also use this feature to sort comments by the length of the thread they are in, in order to find the most popular conversations. It should be pointed out that FutureLearn has recently introduced a facilitator dashboard that identifies the longest threads and most liked comments, and it does it in real time. This makes a significant part of the UoSMOD comments viewer unnecessary. However, it is still highly useful for finding relevant comments retrospectively, as it contains a wider set of filters, such as date, step, and length of the comment.

Other features are more suited for research rather than for practice, such as the correlations viewer. Several researchers have used it for doing post-hoc research about the behaviour of the learners in their courses.

6 Challenges

The development and maintenance of this tool carries several challenges and dependencies. Perhaps the most salient dependency is that of the metadata scraping, as any change in the source site (the FutureLearn admin site) can affect the information retrieval process. For this, a fluent relationship with the platform needs to be maintained so that the developers of the tool are updated about the changes in the platform.

Another challenge is maintaining the privacy and the data protection. For the moment, the whole tool is password protected, but there is information that should not be available to everyone with the password. As the free version of the Shiny framework was used, this did not allow easy compartmentalisation, and a high level of trust had to be placed in all UoSMOD users. This lack of granularity of access in Community version of Shiny may also prevent the combination of data with other institutions.

7 Future work

As future work, there are two actions that are being prioritised. Firstly, a systematic longitudinal use case study is being conducted with different stakeholders such as mentors and educators in the same university where the Dashboard is being developed. These are being exposed to the dashboard, asked to complete a questionnaire, and briefly interviewed before, during, and after they have used it in an instance of a course. The results of the case study will shed light on how learning analytics can make a difference in educational practice. These results will also be used as directions for a third iteration of the UoSMOD development, in which new features will be implemented and others will be modified in response to this second round of feedback from the users.

Another priority is providing different modes of access to different stakeholders, in order to avoid potential data protection risks. For example, mentors of a particular

course should not have access to sensitive information such as the number of upgrades sold in other courses. This will involve replicating the dashboard in bespoke instances.

8 References

1. Reich, J. (2015). Rebooting MOOC research. *Science*, 347(6217), 34-35.
2. Sunar, Ayse Saliha, Nor Aniza Abdullah, Su White, and Hugh Davis. "Personalisation in MOOCs: A Critical Literature Review." In *International Conference on Computer Supported Education*, pp. 152-168. Springer International Publishing, 2015.
3. Cobos, R., Wilde, A., & Zaluska, E. (2017). Comparing attrition prediction in FutureLearn and edX MOOCs. (in press)
4. Clow, D. (2012, April). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134-138). ACM.
5. Cobos, R., Gil, S., Lareo, A., & Vargas, F. A. (2016, April). Open-DLAs: An Open Dashboard for Learning Analytics. In *Proceedings of the Third (2016) ACM Conference on Learning@Scale* (pp. 265-268). ACM.
6. Chitsaz M; Vigentini, L.; Clayphan J. (2016), Toward the development of a dynamic dashboard for FutureLearn MOOCs: insights and directions, in Barker S; Dawson S; Pardo A; Colvin C (eds.), *Toward the development of a dynamic dashboard for FutureLearn MOOCs: insights and directions*, Australasian Society for Computers in Learning in Tertiary Education, Adelaide, presented at Australasian Society for Computers in Learning in Tertiary Education, Adelaide, 27 - 30 November 2016,