FutureLearn data: what we currently have, what we are learning and how it is demonstrating learning in MOOCs.

Lorenzo Vigentini[1], Manuel León Urrutia [2] & Ben Fields[3]

[1] UNSW Sydney, Australia
[2] University of Southampton, UK
[3] FutureLearn, UK
l.vigentini@unsw.edu.au, m.leon-urrutia@soton.ac.uk,
ben.fields@futurelearn.com

ABSTRACT. Compared to other platforms such as Coursera and EdX, Future-Learn is a relatively new player in the MOOC arena and received limited coverage in the Learning Analytics and Educational Data Mining research. Founded by a partnership between the Open University in the UK, the BBC, The British Library and (originally) 12 universities in the UK, FutureLearn has two distinctive features relevant to the way their data is displayed and analyzed: 1) it was designed with a specific educational philosophy in mind which focuses on the social dimension of learning and 2) every learning activity provide opportunities for formal discussion and commenting. This workshop provided an opportunity to invite contributions spanning several areas of investigation and development. The papers collated in this proceeding include: 1) the development of dashboards to support the analytical exploration of FutureLearn Data (León-Urrutia, and Vigentini), 2) the application of analytical methods to understand and improve learners' engagement and participation, especially understanding patterns of communication in FL (Chua, Tagg, Sharples and Rientes) and the comparison of FutureLearn and ED-X data to predict attrition (Cobos, Wilde and Zaluska) ; 3) an example of the use of analytics to support the future pedagogical development of FL MOOCs (Vulic, Chitsaz, Prusty and Ford).

Keywords: MOOCs; visualization dashboard; learning analytics.

1   Introduction

Many higher education institutions have invested in the development of MOOCs. Some have partnered with one or more leading MOOC providers leveraging on the capabilities of different platforms (i.e. Coursera, EdX, FutureLearn etc.) [9, 12]. Others have been experimenting with a collection of open resources and encouraged learners to participate in learning experiences at scale, without the constraints of specific platforms

and promoting a connectivist experience of learning [1, 6, 10, 13, 16, 18]. With the experimentations in learning design, many started to question the effectiveness of the forms of learning that can be supported by the introduction, and given that a large amount of data has become available, it is timely to explore how to best make use of it.

In fact, with the increased availability of MOOC data, there is an opportunity to provide insights to educators and developers into learners' behaviours, and empower learners to understand their patterns of engagement and performance through learning analytics [19]. Providing insights to educators allows exploring learning design at scale and has the potential to inform pedagogy. Empowering learners using learning analytics can improve the learning experience and develop crucial metacognitive skills essential for self-directed and lifelong learners. In more recent times there has been a shift from descriptive analytics to analytics able to inform and direct practice [5, 21, 22]. This was also advocated by Gasevic and colleagues as a key area of further research in their review of research in MOOCs [8]. In fact, despite the existing large body of research, there are two crucial problems hindering the application of learning analytics methods to support and shape pedagogy in MOOCs: 1) the constraints of the platforms (i.e. the tools and course design) and 2) the availability of data when it is needed.

Looking at the wealth of research in MOOCs, a great deal of it is conducted 'post-hoc', when the respective platforms release the data for exploration, and often data is locked within institutions limited by their agreements with platform providers. Research has looked at Coursera data [2, 15] and the dashboard offered to partners' institutions [7]. In addition, the relative openness of EdX, allowed different teams to develop extensions/plugins to access and use analytics [4, 14, 17, 20].

FutureLearn went down a different pathway, focusing on standardization and simplicity, offering data files to partner institutions to enable them to make sense of the interaction occurring in the various courses. Additionally, a report (based on R scripts) is offered to stakeholders, but this is limited in several ways: 1) it is static, 2) it focuses on selected information and, 3) most importantly, it does not provide 'real-time' access to data. The lack of a tool to visualise data from the engagement with FL MOOCs sparked two separate initiatives to develop tools bringing analytics to different stakeholders [3, 11].

2 Scope and opportunity

The workshop was conceived as an opportunity to invite contributions and share the work already done by several FL partner institutions showcasing existing processes, methods and tools used to analyse, present and use the data offered by the FL platform. Submissions were invited along two streams: a research/practitioner track and a technical track. These were intended to present case studies demonstrating how practitioners use the data to inform pedagogical design, what questions and findings researchers uncover in the data (and what is still missing), and the type and nature of technology stack explored to analyse and present data.

The Workshop was intended for those who wish to understand the possibilities offered by the data already offered by FutureLearn, discuss and share innovations, impact

on education, and explore future directions in the application of learning analytics (LA) to Massive Open Online Courses (MOOC) designed and developed in the FutureLEarn platform. It was expected that likely interested participants would be:

- Educators/teachers and researchers,
- technologists and educational developers
- learning scientists and Data scientists/analysts
- academic managers
- entrepreneurs
- and anyone else interested in MOOCs (focusing on FutureLearn in this workshop) and LA.

3 Outcomes of the Workshop

The Workshop was well received, with 32 participants attending. The original aspiration for collating several submissions was well matched by the 3 accepted submissions and the invited papers, which include three broad areas of interest:

1. The development of dashboards to support the analytical exploration of FutureLearn Data (León-Urrutia, and Vigentini),
2. The application of analytical methods to understand and improve learners' engagement and participation, especially understanding patterns of communication in FL (Chua, Tagg, Sharples and Rientes) and the comparison of FutureLearn and ED-X data to predict attrition (Cobos, Wilde and Zaluska);
3. An example of the use of analytics to support the future pedagogical development of FL MOOCs (Vulic, Chitsaz, Prusty and Ford).

More specific details can be found in the various papers, briefly summarised in the following overview.

3.1 The development of dashboards to support the analytical exploration of FutureLearn Data

Two examples of dashboard development were presented, and the code was shared (see end notes). The work done at the University of Southampton and UNSW Sydney followed very similar paths, leveraging a very similar technology stack. Using R scripts, the Shiny dashboards library and a web-server, the two project demonstrated how, starting from the simple data files provided, the data could be processed, visualised and organised for different stakeholders to provide not only an overview of participants' engagement with platform, but allowed a deep drill-down into activities in each individual course providing a wealth of opportunities to initiate conversations with both academic manager and educational development teams behind the design and delivery of FL MOOCs. Both projects were successful in explicitly attempting to fill the analytical gap and make FL data usable for partners. Also, both projects are still ongoing, and new features and improvements are being constantly implemented.

3.2 The application of analytical methods to understand and improve learners' engagement and participation

Another two, very different examples presented how the data provided by FL can be used by researchers to delve deeper into pedagogical questions involving learners engaging with FL MOOCs. The first example (Chua, Tagg, Sharples and Rientes), explores how the engagement with learning conversations evolves in FL MOOCs, keeping into account the pedagogical philosophy behind the 'conversations in context' allowing learners to comment directly in each step (or unit of content). The authors provide a categorization of the learners' contributions quantifying the dynamics of conversations in the discussion activities, detailing how *social learners* contribute in the course steps. The second paper (Wilde, Cobos, and Zaluska) provides a comparison of attrition across two different MOOC platforms. The authors discuss the differences between the datasets provided in Ed-X and FL and applied several machine learning algorithms on the data to predict attrition levels for each course. The analysis suggests that the attribute selection must be considered carefully in each scenario as their analysis identified different patterns of outcomes using the same predicting algorithms.

3.3 The use of analytics to support the future pedagogical development of FL MOOCs

The final paper takes a practitioner perspective and shows how the insights emerging from the data provided through the UNSW dashboard has been used to directly engage the academic leas into a conversation about the pedagogy and effectiveness of the course. The conversation led to refinements of the Engineering MOOC which ultimately led to improvements in the student experience (Ford et al. in press).

3.4 Overall takeaways

As well as the contributions from the speakers, several participants representing four continents brought to the table questions and issues they face, the comparison with other MOOC platforms, highlighting the strengths and weaknesses of the FL learning analytics offering and shared their own experiences.

In line with the expectations, the workshop provided a tangible opportunity to:

— Get an idea of the state of the art of work with FutureLearn data across institutions, disciplines and roles;
— Discuss cases, issues and problems, sharing outcomes (both successes and failures in using the data offered);
— Reflect on the impact of the work presented on learning design and the learners' experiences;
— Enable the development of common tools that educators and researchers may be able to re-use in their own contexts;
— Connect people with one another, in the broad area of data and LA applied to MOOCs and FutureLearn in particular.

— Explore opportunities of sharing results for cross-course analysis and benchmarking.

4   Future Directions

As a growing company, FutureLean has demonstrated their commitment to support partners, collaborate and co-develop effective solutions to improve research opportunities, learning design and ultimately the learners' experience.

However, several limitations were discussed, particularly by partners currently delivering (or developing) courses in FL. Here are some issues worth highlighting:

— **Simplicity of datasets does not equate to accessibility of data**: despite choosing a simple set of core datasets, many noted the gap (currently filled by the work presented with dashboards) in making the data usable by partners.
— **Controlling the data is not always a good thing**: especially when FL has relatively limited analytical capacity, it would be good to allow partnerships to support the understanding of engagement in FL and how this differs from other platforms. In this sense, providing similar datasets to other platform will help to clearly determine the value added of the FL pedagogical design model.
— **Personal information and data triangulation**: this was seen as a major drawback for all partners present. In order to extract meaningful interpretations and allow for data triangulation, support interventions and post-course conversion, the current approach to data privacy adopted by FL is perceived as meaningless to Universities and educational organisations which are already well versed with the management and governance of students' personal details.
— **FLAN may not be enough**: Most of the FutureLearn Academic Network events take place in the UK, with a few exceptions in other European countries. This was the first workshop of its kind at LAK. Most participants saw this as an excellent opportunity to present and share the work done by partners outside the UK/EU context and increase the opportunities for potential collaborations internationally. Given the expansion of FL partners in both Asia and the US, this was a welcome event which it is worth continuing in the future.

5   References

1. Arnold, P. et al. 2014. Offering cMOOCs collaboratively: The COER13 experience from the convenors' perspective. eLeanrning Papers. 37, (2014), 63–68.
2. Chandrasekaran, M.K. et al. 2015. Learning Instructor Intervention from MOOC Forums: Early Results and Issues. arXiv:1504.07206 [cs]. (Apr. 2015).
3. Chitsaz, M. et al. 2016. Toward the development of a dynamic dashboard for FutureLearn MOOCs: insights and directions. Proceeding of ASCILITE (Adelaide, SA, 2016).
4. Cobos, R. et al. 2016. Open-DLAs: An Open Dashboard for Learning Analytics. Proceedings of the Third (2016) ACM Conference on Learning @ Scale (New York, NY, USA, 2016), 265–268.
5. Corrin, L. et al. 2015. Loop: A learning analytics tool to provide teachers with useful data visualisations. Proceedings of ASCILITE (Perth, Australia, 2015), 409–413.

http://www.2015conference.ascilite.org/wp-content/uploads/2015/11/ascilite-2015-proceedings.pdf

6. Dawson, S. et al. 2015. Recognising learner autonomy: Lessons and reflections from a joint x/c MOOC. Proceedings of Higher Education Research and Development Society of Australia 2015.

7. Do, C.B. et al. 2013. Self-Driven Mastery in Massive Open Online Courses. MOOCs FORUM. 1, P (Sep. 2013), 14–16.

8. Gasevic, D. et al. 2014. Where is research on massive open online courses headed? A data analysis of the MOOC Research Initiative. The International Review of Research in Open and Distributed Learning. 15, 5, 134-176 (Oct. 2014).

9. Jordan, K. 2014. Initial trends in enrolment and completion of massive open online courses. The International Review of Research in Open and Distributed Learning. 15, 1, 134-60.

10. Kanwar, A. and Mishra, S. 2015. The impact of OER and MOOCs on ODL: an international perspective. International Distance Education Development Forum, Peking University, Beijing, China, 10 October 2015. http://oasis.col.org/handle/11599/1734

11. Leon Urrutia, M. et al. 2016. Visualising the MOOC experience: a dynamic MOOC dashboard built through institutional collaboration. Proceedings of the European MOOC Stakeholder Summit 2016. 461-471.

12. McAuley, A. et al. 2010. The MOOC model for digital practice. (2010). http://www.davecormier.com/edblog/wp-content/uploads/MOOC_Final.pdf

13. McIntyre, S. et al. 2015. Learning to Teach Online-Evolving approaches to professional development for global reach and impact. e-Learning Excellence Awards 2015: An Anthology of Case Histories. 128–140.

14. Pardos, Z.A. and Kao, K. 2015. moocRP: An Open-source Analytics Platform. Proceedings of the Second (2015) ACM Conference on Learning @ Scale (New York), 103–110.

15. Parod, B. 2014. Developing an Analytics Dashboard for Coursera MOOC Discussion Forums. https://www.cni.org/wp-content/uploads/2014/12/Mon_Parod_DevMOOCDashboard.pdf

16. Rosé, C.P. et al. 2015. Challenges and opportunities of dual-layer MOOCs: Reflections from an edX deployment study. Proceedings of the 11th International Conference on Computer Supported Collaborative Learning (CSCL 2015). 848-851.

17. Ruiz, J.S. et al. 2014. Towards the Development of a Learning Analytics Extension in Open edX. Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality (New York, NY, USA), 299–306.

18. dos Santos, A.I. et al. 2016. Opportunities and challenges for the future of MOOCs and open education in Europe.

19. Siemens, G. et al. 2011. Open Learning Analytics: an integrated & modularized platform. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques..

20. Veeramachaneni, K. et al. 2014. MOOCdb: Developing Standards and Systems to Support MOOC Data Science. arXiv:1406.2015 [cs]. (Jun. 2014).

21. Wise, A.F. 2014. Designing Pedagogical Interventions to Support Student Use of Learning Analytics. Proceedings of the Fourth International Conference on Learning Analytics And Knowledge (New York, NY, USA, 2014), 203–211.

22. Wise, A.F. et al. 2016. Developing Learning Analytics Design Knowledge in the "Middle Space": The Student Tuning Model and Align Design Framework for Learning Analytics Use. Online Learning. 20, 2. http://olj.onlinelearningconsortium.org/index.php/olj/article/view/783