Challenge and Potential of Fine Grain, Cross-Institutional Learning Data

Alan Dix
Talis
48 Frederick Street
Birmingham, UK, B1 3HN
and
HCI Centre
School of Computer Science
University of Birmingham
Edgbaston
Birmingham, UK, B15 2TT

ajd@talis.com

http://alandix.com/academic/papers/LS2016/

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
Copyright is held by the owner/author(s).
L@S 2016, April 25-26, 2016, Edinburgh, Scotland UK
ACM 978-1-4503-3726-7/16/04.
http://dx.doi.org/10.1145/2876034.2893429

Abstract
While MOOCs and other forms of large-scale learning are of growing importance, the vast majority of tertiary students still study in traditional face-to-face settings. This paper examines some of the challenges in attempting to apply the benefits of large-scale learning to these settings, building on a growing repository of cross-institutional data.

Author Keywords
learning analytics; education technology; reading lists; MOOCs; OER; linked data.

Introduction
Can the benefits of scale work for traditional face-to-face institutions or smaller distance classes as well as MOOCs and similar environments? While large-scale courses can attract tens of thousands of students, the majority of the nearly 100 million new tertiary student enrolments worldwide [9] are in traditional face-to-face settings and most of the rest in much smaller distance courses. This is large-scale by any definition, but especially vast in terms of heterogeneity.

The number of tertiary students is expected to grow from around 180 million to 260 million by 2025 ([3] as reported in [6]). Even if many are in large-scale courses (which is unlikely), it is clear that we need to

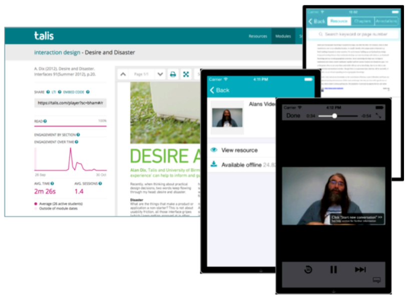

Figure 1. Universal media player, web and mobile version.

Figure 2. Document learning analytics

use the lessons learnt from large-scale learning to benefit the hundreds of millions in traditional settings.

Potential

Talis provides teaching and learning software for higher education including a reading list management module used by 86 institutions in 8 countries, including over 50% of UK universities. Currently, Talis serves nearly half a million lists with over twenty million references.

While the number of students covered by these is tiny compared to the world statistics, it is large enough to start to consider the potential for using the techniques of large-scale data. For example, it would be possible to offer textbook recommendations to educators based on usage data from other similar educators.

In addition, Talis is developing a universal media player currently under pilot in a project called Lighthouse (fig. 1). This player allows fine-grained tutor and student annotation of videos, PDF documents, etc., and also usage tracking at the level of page views within a PDF or sections within a video, not simply click-through (fig. 2). That is as well as a large volume of students, there will be highly detailed information available about each. In principle this could allow, for example, sharing of annotations between students studying the same material in different institutions, allowing new forms of peer interaction.

Challenges

However, despite the clear potential of this combination of large and fine scale, there are challenges in leveraging this volume of data to aid individual academics and students.

- *Heterogeneity of courses* – There are many different kinds of courses both within institutions and between, as well as very different teaching styles. In addition, individuals differ markedly.

- *Individuality* – In step-by-step programmed instruction, progress is easy to measure, but things are far more complex where teaching is less constrained and when individuals' patterns of reading/viewing media vary.

- *Cross-institutional issues* – When looking at data across institutions heterogeneity is increased, and additional factors arise, not least connected with privacy and ownership.

MOOCs avoid many of these challenges. By having a large number of students addressing the same material provided by the same organization it is possible to perform large-scale analysis on relatively homogeneous data and also more easily enable pedagogic elements such as peer learning.

We now look at each challenge in a little more detail.

Heterogeneity of courses

This is an issue within a single institution with many different kinds of material from lab-based work to reading focused topics, but in addition institutions vary.

Looking simply at reading list length, these range from over a thousand items for a single module to some with just a core textbook and a few additional resources. It would be tempting to see this in terms of broad subject areas such as humanities (long list) vs. sciences (core textbook). However, this would be simplistic because

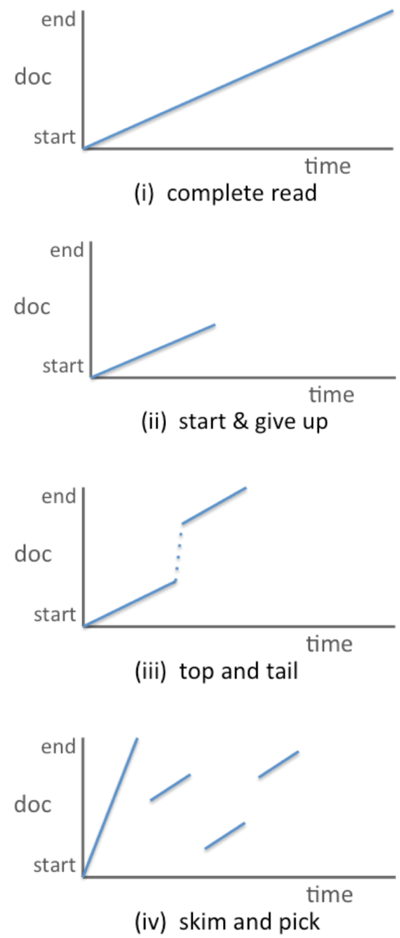

Figure 3. Patterns of document access (N.B. possible examples, not based on data analysis)

of variations within areas (e.g. a creative writing course may be 'hands on' rather than reading based) and radically different ways of teaching the same topic (e.g. theory-then-apply vs. problem-based learning). More content-based approaches are needed.

The Open University (OU) analysed module content in terms of seven kinds of activities (assimilative, finding information, communication, productive, experiential, interactive, assessment) and then looked at the relative proportions of these and created clusters [8]. The clusters turned out not to be subject based, but rather related to different teaching styles: constructivist, assessment-driven, balanced-variety, and social constructivist. Having done this they were then able to correlate these with student experience measures.

This OU analysis seems to require an initial hand categorization of activities; however, we envisage that similar analysis may be possible based on content alone, especially given different kinds of content (video, article, book, quiz).

Individuality

While the content of MOOCs is identical for large numbers of students, the students themselves vary; of course this is even more challenging when combined with heterogeneity of courses. Learning styles, of multiple kinds, have become a minor education 'industry'. While there is evidence of differences, empirical evidence of clear pedagogic value is still scant compared to other individual factors [2].

There are successful examples grouping/clustering students automatically. P2PU (Peer-to-Peer University) have used simple user profiling (principally time zone) to organize students into smaller groups [5], and Purdue 'Signals' is based on a combination of historical student data and in-course behaviour [1].

Fine-grained behavioural data offers more potential for understanding individual behaviour, but is also correspondingly more complex. The engagement graph in figure 2 shows that although a few students have read all the document, many have just read the first part; this was used by the tutor to suggest to students that even if they skip the middle of the document they should read the last section [4].

However, aggregation inevitably hides individual differences. While it would be possible to view each individual, it would also be useful to be able to classify student reading behaviours, rather like the OU have done for their modules. We have not done this yet, but fig. 3 shows the kinds of reading patterns we might expect to see.

Cross-institutional issues

Perhaps most difficult are the challenges faced because data arises from multiple institutions. This complicates both the preceding challenges, but also adds issues of privacy and ownership.

Talis encourages customers to make reading lists available with open data licences and the reading list software makes the data available in RDF linked data format, providing the beginnings of an 'education graph' linking resources across institutions [7].

However, this is content only; student usage, progress, etc., is confidential and private to the institution.

There would be potential benefits to institutions if suitably anonymised student data were combined, analysed, and the results fed back to improve individual student experience. However, not all institutions may wish to share such data out of concern for privacy of students or perceived value of the data. Furthermore, the balance of control and ownership between the institution and the student may vary depending on each institutions' policies, student agreements, and national laws. These legal and ethical problems almost certainly outweigh technical challenges.

Summary
This paper has discussed a number of challenges if large-scale learning benefits are to be brought to bear on traditional face-to-face institutions. These have been driven by the particular data available and systems developed at Talis, and our work on addressing the challenges, but the lessons are likely to apply equally to providers of VLEs and other widely used educational software.

References

1. Arnold, K. and Pistilli, M. 2012. Course signals at Purdue: using learning analytics to increase student success. In *Proc. of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)*. ACM, NY, USA, 267-270. doi:10.1145/2330601.2330666
2. Coffield, F., Moseley, D., Hall, E. and Ecclestone, K. (2004) *Learning Styles and Pedagogy in post-16 learning: a systematic and critical review*. London: Learning and Skills Research Centre http://www.lsneducation.org.uk/research/reports/
3. Davis , D. and Mackintosh, B. (eds). *Making a Difference: Australian International Education*. UNSW Press (2012)
4. Dix, A. More than one way to flip a class: learning analytics for mixed models of learning. *APT 2015*, Greenwich, 7th July 2015. http://alandix.com/ academic/papers/apt2015-more-than-one-way/
5. Ewald, C. (2013). *The Mechanical MOOC – Behind the Machine*. P2PU, 15th Jan. 2013. http://info.p2pu.org/2013/01/15/the-mechanical-mooc-behind-the-machine/
6. Gibney, E. A Different World. *Times Higher Education*, January 31, 2013. https://www.timeshighereducation.com/features/a-different-world/2001128.article
7. Heath, T., Singer, R., Shabir, N., Clarke, C., and Leavesley, J. Assembling and Applying an Education Graph based on Learning Resources in Universities. *Proc. of the 2nd International Workshop on Learning and Education with the Web of Data (LiLe-2012)*, 2012. http://ceur-ws.org/Vol-840/
8. Rienties, B. Leveraging the full value of learning analytics: The power of learning design. (keynote) *LACE SoLAR Flare 2015*, OU, Milton Keynes, UK (2015). http://www.slideshare.net/laceproject/ leveraging-the-full-value-of-learning-analytics-the-power-of-learning-design
9. UNESCO. *Education: Gross enrolment ratio by level of education*, UNESCO Institute of Statistics (accessed 7/1/2016) http://data.uis.unesco.org/ index.aspx?queryid=142