

Learning Analytics: From Big Data to Meaningful Data

Agathe Merceron

Beuth University of Applied Sciences, Berlin, Germany
merceron@beuth-Hochschule.de

Paulo Blikstein

Stanford University, California, USA
paulob@stanford.edu

George Siemens

University of Texas at Arlington, USA
gsiemens@gmail.com

ABSTRACT: This article introduces the special issue from the 2015 Learning Analytics and Knowledge conference. We describe the current state of the field and identify some of the trends in recent research. As the field continues to expand, there seem to be at least three directions of vigorous growth: 1) the inclusion of multimodal data (gesture, eye-tracking, biosensors), 2) the diversification of learning environments (MOOCs, classrooms, hands-on learning), and 3) new types of research questions considering a broader set of learning-related constructs (e.g., moving away from the focus on student retention).

Keywords: LAK '15, 5th International Learning Analytics and Knowledge Conference, special issue.

1 Introduction¹

The history of technology is full of examples of automation techniques that had revolutionary impact on various human activities. Tasks done laboriously by hand were suddenly taken over by machines with gains of productivity of one or more orders of magnitude. In education, researchers and policymakers have desired the Holy Grail of productivity for decades. Even the early behaviourists expressed their wish for machines that would teach and assess with very little human intervention (Skinner, 1968). The relatively new field of learning analytics and educational data mining has revived our hopes of increased effectiveness in the educational field. In the learning analytics community, the research started to have more exposure after high profile initiatives gained traction: massive online courses, online video-based learning, educational apps, and the massive availability of all sorts of computing devices that are not just personal but portable, providing access to increased amounts of user data in electronic form. It was not

¹ LAK '15, the 5th International Conference on Learning Analytics and Knowledge, took place at Marist College, Poughkeepsie, New York, from March 16 to 20, 2015 under the motto: "Scaling Up: Big Data to Big Impact."

(2015). Learning analytics: From big data to meaningful data. *Journal of Learning Analytics*, 2(3), 4–8. <http://dx.doi.org/10.18608/jla.2015.23.2>

uncommon in learning analytics gatherings a few years ago to hear computer science researchers claim that in this new golden age of analytics, theories about learning were not necessary, because the theories would emerge from the ocean of data that had just been made available from log files, clickstreams, and eye trackers. Through the lens of 2016, however, the focus on LA is more holistic and integrative, accounting for social and affective learning practices based on existing research within the learning sciences (Calvo & D’Mello, 2010).

In its early years, the learning analytics community was, as any early stage of field building should be, an attempt to find coherence; in 2015, coherence has certainly appeared. This evolution shows some interesting trends. From simply counting events in clickstreams and reporting their percentages we moved to trying to establish correlations, and then finally to develop causal explanations and theoretical models. We went from a relatively naïve project of “automation of assessment” to a much more complex and challenging endeavor where humans and machines work together, and analytics have become more of a source of information for educators to make complex decisions that cannot be outsourced to machines. A considerable part of our community had a utopian (or dystopian) vision of a future in which each student would have a personalized playlist or intelligent system that would make education happen massively and at a very low cost. Now, it seems that our vision is much more realistic and modest; changes in education are not seen as sudden and revolutionary, but more incremental and difficult.

It would appear that the learning analytics community is becoming more focused on broad research from many data sources and targeting many nuanced questions about what it can deliver. In the 5th International Conference on Learning Analytics and Knowledge (LAK ’15), we witnessed many emerging trends and concerns, a community that is becoming more mature, more aware of its limitations, and yet more focused on its ambitions. One of the emerging major trends is expanding the sources of data from which learning analytics can be conducted. Previous conferences saw many sessions concerned with learning management system (LMS) data at the classroom level; this year, we had one session. Instead, participants were gathering and analyzing data from a host of environments: MOOCs, classrooms, and face-to-face hands-on learning environments.

Another trend relates to the types of data researchers are using. We are more and more aware that teaching and learning are multimodal processes, which include voice, gesture, joint visual attention, and several different biological and mental processes happening at the same time. Capturing all of that using a single modality is not sufficient, especially for face-to-face learning environments. One session featured the word “multimodal,” but the emphasis and importance of multimodal data was observed across a range of sessions and was highlighted in the final keynote address.

Another major trend was in the depth of the analysis undertaken. As one can reasonably expect, the community has made significant headway in the algorithms available for analyzing learning data. In particular, researchers are able to examine discourse using greater semantic understanding, garner more in-depth insights from network analysis, and more accurately triangulate across a broad set of modalities.

(2015). Learning analytics: From big data to meaningful data. *Journal of Learning Analytics*, 2(3), 4–8. <http://dx.doi.org/10.18608/jla.2015.23.2>

The final trend to note is the increasing complexity and applicability of research findings. Increased complexity points to the growing set of dependent variables being analyzed. Instead of merely being concerned with student retention and traditional notions of achievement, researchers are beginning to consider a broader set of learning-related constructs while also paying additional attention to at-risk populations. Within this increased complexity, in terms of what and whom researchers are studying, we see an increase in applicability. Over the years, the research has moved ever closer to the various participants in day-to-day teaching and learning practices.

2 SCALING UP

The theme for the LAK '15 conference — “Scaling Up: Big Data to Big Impact” — reflected the evolution of the field. But what is scaling up? What do we understand by “big impact”? Clearly, the papers and discussions at the LAK '15 conference provide sound evidence that the LA community is scaling up, and moving from simply “big” data to “meaningful” data. The program committee has broadened to include related fields and better reflect our growing community. The number of submissions increases every year and was the biggest since LAK was established in 2011. The number of high quality submissions was substantial and, consequently, was reflected in the accepted papers. As well, another measure of “scaling up” is the number of attendees: LAK '15 was the biggest and most diverse conference to date.

The format of the conference has been expanded to include a practitioner track that intertwines with the research track, allowing greater opportunities for cross-fertilization between research and practice. The size of the pre-conference workshops, tutorials, and doctoral consortium also increased. The growth of the doctoral consortium is particularly exciting, as it brings new ideas that will continue to shape the learning analytics community, challenging the assumptions that undergird our work and how we interpret research findings. At the same time, it creates a space for students to observe and contribute to the development of the field, enabling them to gain familiarity with the rich history of pre-LAK learning analytics that helped to motivate the creation of this field. As the emergent researchers, doctoral students will eventually infuse learning analytics into the fabric of their institutions and so are crucial to the idea of scaling up learning analytics in the long term.

Within learning analytics, Big Data takes on a number of different instantiations. First, the field is branching out to a wider set of data sources and modalities, which is essential in conducting research that has Big Impact. At the same time, these new data sources and learning platforms are fuelling opportunities for the community to develop new analytic techniques, an important next step for driving our level of impact. Finally, an important consideration is the mounting concern around user privacy issues. If Big Data is to have Big Impact in education, addressing privacy concerns will be critical. Generating sufficiently generalizable and robust results must warrant the claims and interventions that we propose through our research. Hence, the challenges and opportunities in Big Data venture far beyond mere terabytes of

(2015). Learning analytics: From big data to meaningful data. *Journal of Learning Analytics*, 2(3), 4–8. <http://dx.doi.org/10.18608/jla.2015.23.2>

information into issues of data access, privacy, reliability, and generalizability.

3 THIS ISSUE

“Big Impact” has many different facets. This special issue illustrates four different perspectives in extended articles from papers presented at LAK ’15, each undergoing additional peer review. The first, by Martinez-Maldonado, Pardo, Mirriahi, Yacef, Kay, and Clayphan, concerns building helpful and intuitive Learning Analytics user interfaces using the LATUX workflow. The paper outlines how to integrate methods from software engineering and human–computer interaction with pedagogical requirements to develop appropriate visualizations. A case study of a dashboard to support instructors while groups of students learn on an interactive tabletop illustrates the approach.

The second paper, by Snow, Allen, Jacovina, Crossley, Perret, and McNamara, concerns the investigation of novel analysis methods. The article proposes merging entropy with natural language processing methods to detect flexibility in student essays. Their results indicated that “the relation between students’ flexibility in writing style and their prior literacy skills can be detected reliably after only a few essays” (p. 48).

The third perspective, that of Ferguson and Clow, considered the importance of undertaking replication studies for verifying the generalizability of published results. They describe their attempt to replicate a 2013 study by Kizilcec, Piech, and Schneider identifying four clusters of student behaviours in MOOCs offered on the Coursera platform. The replication uses data from four MOOCs “that employ social constructivist pedagogy.” It turns out that Ferguson and Dow could only put in evidence one group described in the previous study: the “Completing” group. Fine-tuning their analysis, Ferguson and Dow found seven typical behaviours. These patterns remained stable in a subsequent repetition of two of these MOOCs.

Finally, Kovanović, Gašević, Dawson, Joksimović, Baker, and Hatala examine distilling and establishing good analysis practices. Their paper reviews different methods that researchers have used to estimate time-on-task, focusing on learning management systems. Kovanović et al. demonstrate how different methods and interpretations used to describe “time-on-task or time online” can effect results. Their findings stress the importance of clearly formulated hypotheses and fully describing how such dimensions (e.g., time) are defined, measured, and analyzed.

Over the coming years, we see the convergence of these perspectives into a broader model of learning analytics research. For example, one could imagine work that leverages LATUX to visualize the behavioural patterns of MOOC users, or essay writers. Similarly, one could see a clear intersection between time-on-task estimations and work conducted on essay writing or in a massive course. As we continue to grow the field, we must be conscious about the practices being established, and then build upon and question that

(2015). Learning analytics: From big data to meaningful data. *Journal of Learning Analytics*, 2(3), 4–8. <http://dx.doi.org/10.18608/jla.2015.23.2>

work, maintaining a high quality of scholarship and optimizing the validity of this research.

4 CONCLUSION

The learning analytics community, against the backdrop of the very successful past five years, now has a considerable challenge before it. We are past the days of exaggerated excitement over massive online courses, learning management systems, iPads in classrooms, miraculous teaching software, or magical learning recommendation systems that know exactly what you don't know and need to learn. We are entering a phase of increasing complexity with a constantly shifting education ecosystem. However, amidst these shifts lies a prime opportunity for learning analytics to demonstrate its utility in ways that are meaningful and impactful to learning. Realizing such an impact will require increased coordination and collaboration in the learning analytics community, and with related communities in educational data mining, artificial intelligence, and the learning sciences, among others. We sit at the cusp of change, and within this natural evolution, a large portion of that success can be detected in the work presented at the 2015 Learning Analytics and Knowledge Conference.

REFERENCES

- Calvo, R., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1), 18–37.
<http://dx.doi.org/10.1109/T-AFFC.2010.1>
- Kizilcec, R., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the 3rd International Learning Analytics and Knowledge Conference (LAK '13)*, 170–179.
<http://dx.doi.org/10.1145/2460296.2460330>
- Skinner, B. (1968). *The Technology of teaching*. New York: Appleton-Century-Crofts.