

Research with Simulated Data

Arnon HersHKovitz, Simon Knight, Jelena Jovanović, Shane Dawson, Dragan Gašević

Editors, Journal of Learning Analytics

jla.editorial@gmail.com

ABSTRACT: This first issue of the Journal of Learning Analytics in 2017 features a special section of invited papers from the recent Learning Analytics and Knowledge conference (LAK'16). The theme of the conference, and this special section, relates to the need for Learning Analytics research to challenge our methodological and theoretical assumptions and build new interdisciplinary connections to further our thinking.

Keywords: learning analytics, LAK, simulated data, simulated learners

1 INTRODUCTION

A constant challenge in learning analytics research is the capacity to obtain large data sets. There is presently a vast number of emerging research that provide critical insight into learning processes developed from a course or program of study. Access to institutional wide data is growing, however these data sets are frequently at a level divorced from the context in which the learning behavior was derived. We have no doubt that as technology adoption further increases and the models of teaching practice further evolve to leverage the affordances such technologies can bring to education there will be a parallel increase in the amount and granularity of learner data. On the one hand, this data provides an essential ingredient in our pursuit for understanding student learning and the development of more personalized learning practices. On the other hand, this approach raises important ethical questions that still need to be adequately addressed. The first paper in the special section begins to unpack this tension. However, an alternate approach may also involve the generation of simulated data sets.

The Second Workshop on Simulated Learners was held in conjunction with AIED 2015, and in its description, the organizers stated that:

As learning environments become increasingly complex and are used by growing numbers of learners [...] the need for simulated learners [...] is compelling, not only to enhance these environments with artificial agents, but also to explore issues using simulation that would be otherwise be too expensive, too time consuming, or even impossible using human subjects ¹.

¹ Quoted from the workshop website, <https://sites.google.com/site/simulatedlearners> [accessed on January 2017].

(2017). Research with simulated data. *Journal of Learning Analytics*, 4(1), 1–2. <http://dx.doi.org/10.18608/jla.2017.41.1>

Indeed, some papers presented to the LA community have already based their analysis on simulated learners (e.g., Pelánek, Rihák, & Papoušek, 2016). Other proponents of using simulated data claim that "the application of synthetic data will not only accelerate the creation of complex and layered learning analytics infrastructure, but will also help to address the ethical and privacy risks involved during service development" (Berg, Mol, Kismihók, & Sclater, 2016, p. 107). The use of simulated data sets has both practical and ethical implications.

The use of simulated learners however, represents a deviation in research for learning analytics. The process provides clear practical benefits, however, there is a blurring in the relationship between simulated and real as the findings are interpreted, implemented and actioned. When applying simulated data, it is important to ensure the methodology of constructing the "learner" population is clearly presented and that findings are carefully, honestly examined in light of that use. This is relevant to many (if not most of the) studies undertaken under the big umbrella of "learning analytics".

This editorial is not to cast a shadow on the use of simulated data nor to actively promote its affordances and benefits. The editorial serves as a reminder to all learning analytics researchers and advocates that the use of any data must be carefully considered, rigorously analysed, interpreted through a strong theoretical framework and actioned with care and caution.

2 IN THIS ISSUE

This issue consists of a Special Section presenting five invited papers from the recent LAK'16 that was held in Edinburg, UK (April 25-29). This section is guest-edited by the Conference Program Chairs, Shane Dawson (University of South Australia), Hendrik Drachler (Open University of the Netherlands), and Carolyn Penstein Rosé (Carnegie Mellon University). The papers in this section discuss various topics—including interfaces, writing analysis, role modelling, learners' dashboard, and the cross-overs between humans and machines—thus demonstrating the wide range of topics presented in the conference and the rainbow of ideas shared throughout the learning analytics community. Furthermore, the presented studies involve both school and university populations, as well as university-level MOOC learners, which well reflects the varied target audience the LA community has aimed at.

More information about the Conference and about the selected papers are reported in the Special Section Editorial. We hope you will enjoy this issue.

REFERENCES

- Berg, A.M., Mol, S.T., Kismihók, G., & Sclater, N. (2016). The role of reference synthetic data generator within the field of learning analytics. *Journal of Learning Analytics*, 3(1), 107-128. <http://dx.doi.org/10.18608/jla.2016.31.7>
- Pelánek, R., Řihák, J., & Papoušek, J. (2016). Impact of data collection on interpretation and evaluation of student models. *Proceedings of the Sixth International Learning Analytics & Knowledge Conference (LAK'16)*, 40-47. <http://dx.doi.org/10.1145/2883851.2883868>