Current state of Learning Management Systems' log data-based learning analytics

Ángel Hernández-García, Universidad Politécnica de Madrid, ORCID: 0000-0002-6549-9549

Emiliano Acquila-Natale, Universidad Politécnica de Madrid, ORCID: 0000-0003-2164-8386

Julián Chaparro-Peláez, Universidad Politécnica de Madrid, ORCID: 0000-0002-0920-4005

Abstract

The application of learning analytics techniques to log data from Learning Management Systems (LMS) has raised increasing interest in the past years. Advances in this field include the selection of adequate indicators and development of research frameworks. However, log data-based analysis of courses still poses some obstacles and challenges for researchers and practitioners in order to effectively improve and optimize learning processes. This paper highlights the challenges, and presents approaches that can help complement log data-based learning analytics. These approaches may be especially effective in collaborative settings, and include analysis of information flows, social interactions, and content analysis. This conceptual work aims to promote the debate surrounding the need for comprehensive and comparable studies and frameworks, and to foster advances in log data-based learning analytics.

Keywords

Learning Analytics, Educational technology, Learning Technologies, E-learning/Online learning, Learning and Instructional Technology, Technology

Introduction

The growing use of Learning Management Systems (LMS), advances in statistical analysis, the rapid development of software and analytics methods, and the increasing interest in the education field to apply the principles of business analytics to learning processes, have led to the emergence of educational data-mining and learning analytics as one of the most promising research fields in computer-supported education.

The main principle of learning analytics lies on the extraction of *useful* and *actionable* information from the large amount of data generated in online learning systems–i.e. LMS log systems–to inform the different learning actors–institutions, instructors and students–in order to improve learning processes.

While the objective of the application of learning analytics techniques may vary from case to case–prediction of academic success, implementation of early-warning systems, reduction of attrition rates, etc.– the analysis mostly relies on one source: the data stored on LMS logs.

Approaches to log data-based learning analytics

The main challenge is to decide which data can provide useful information or how to aggregate and present data in a format that may offer any additional value to the different actors. Agudo-Peregrina et al. (2013) define interactions and their representation as data log records as the basic contextualized data units needed for learning analytics.

Initial studies on interactions included very basic indicators, such as the number of logins in the system. As the interest on the study of LMS log data increased, researchers start to add indicators for exploratory research, such as number of views of specific course elements, discussion views, forum posts or time spent in the LMS (e.g., McFadyen & Dawson, 2010).

While this is a valid approach, and still an extended practice among researchers, it is also very problematic because of the lack of criteria for selection of indicators. Agudo-Peregrina et al. (2013) address this problem with their proposal of a system-independent classification of interactions. Their framework defines three types of classifications: agent-based (student-student, student-teacher, student-content and student-system interactions), frequency-based (with three different levels of use) and participation mode-based (active and passive interactions). The authors analyse interactions from Moodle logs in face-to-face and online courses, and argue that such a classification–after adaptation to the particular cases of each LMS log system–may help unifying and generalizing research results for theory building. Ongoing research (e.g. Joksimović et al, 2015) also builds on Agudo-Peregrina et al. (2013), focusing mainly on agent-based classifications. These studies support the use of indicators from logs–e.g. student-student interactions–as predictors of final grade in online learning but not in LMS-supported face-to-face courses.

Problems and challenges

Systematic classification of interactions, however, does not completely solve the problem. In fact, the disparity of results from different studies using either an arbitrary selection of indicators or systematic classifications point out the importance of contextual factors–instructional conditions–and suggest the need for theory-driven research when using log data for learning analytics (Gasevic et al., 2016). Furthermore, two additional, less studied problems, also need to be addressed in log data-based learning analytics of online courses: data dimensionality and accuracy of measurement.

Data dimensionality is especially relevant in online computer-supported collaborative learning. So far, log data-based learning analytics mostly focuses on one dimension or specific aspect of data: frequency of interactions, at individual or course levels. Such a perspective leaves out essential information about collaborative learning processes, where social construction of learning is expected to happen. This essential information relates to three main aspects of learning–information flow, social interactions, and meaning of interactions–that require enriching original log data with additional information:

- Regarding flow of information, log data should include not register the interaction, but also its temporal distance to related interactions–e.g. time between responses to messages.

- Concerning social interactions, information about the source and target of the interaction needs to be included. Approaches using SNA provide additional predicting variables related to centrality measures and easy-to-understand data visualizations (Hernández-García et al., 2014).
- Finally, raw data do not provide direct information about the meaning or quality of interactions. Natural language processing techniques and discourse analytics would greatly help researchers to differentiate between "low-quality" and "high-quality" interactions when they conduct learning analytics from LMS log data.

Accuracy of measurement is still a pending issue that refers to the inference of log data that does not represent the variable under study accurately, or that represents data involving logged and non-logged data. Good examples are the time spent on the LMS, on an activity or reading course materials. The usual approach in this case is to register the time between two consecutive interactions in the LMS, and to assign a value of the difference to the first interaction. However, this value does not take into account whether the student or instructor has actually been performing the action during the whole time between both interactions. A similar principle applies to the estimation of time spent reading course materials, especially when they are available for download and offline reading.

References

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior, 31*, 542–550.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet and Higher Education, 28*, 68–84.

Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. *Computers and Education, 87*, 204–217.

Hernández-García, Á., González-González, I., Jiménez-Zarco, A. I., & Chaparro-Peláez, J. (2014). Applying social learning analytics to message boards in online distance learning: A case study. *Computers in Human Behavior, 47*, 68–80.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588–599.