

LEARNING ANALYTICS IN PRACTICE: SETTING UP A LABORATORY FOR ACTION RESEARCH AT THE UNIVERSITAT OBERTA DE CATALUNYA

J. Minguillón, Universitat Oberta de Catalunya, C. Appel, Universitat Oberta de Catalunya, F. Santanach, Universitat Oberta de Catalunya

Introduction

Since its inception in 1994 as a purely online university, the Universitat Oberta de Catalunya (UOC) has been able to position itself among the main universities of the Catalan and Spanish university systems. Most of the students at the UOC (currently more than 60,000) are adults who have a profile that could hardly fit into the traditional university system, thus finding in the UOC an opportunity to start or continue their higher education grades, in a very innovative environment. The intensive use of ICT for both the teaching/learning processes and management allows researchers and practitioners to obtain data about what takes place in the UOC Virtual Campus, which is continuously being improved according to such findings.

The e-Learn Center – the UOC's e-learning research and innovation institute - is the element that the university avails of to facilitate applied research in e-learning, to foster innovation in this area and to convey both internally and externally that e-Learning is a constituting element of the university's identity. The role of the eLearn Center in UOC is two-fold. Internally, it plays an active role in ensuring that research, innovation and practices at UOC are constantly feeding into each other, with the purpose of improving the quality of the students' experience. Externally, it seeks to reaffirm and improve UOC's position in the international arena by strengthening international contacts, supporting the publishing and dissemination of applied research and innovation results and making the UOC environment a place of interest for researchers.

Besides virtual learning environments, there are external variables that have also affected the way students act and interact with the UOC. These include the creation of new degrees according to the Bologna model, which introduced the concept of competence but without resolving its connection with traditional learning resources, and the economic crisis that has caused modifications in the enrolment patterns of the students (currently now they are enrolling in fewer courses each semester). In addition, the increased use of mobile devices, especially smartphones, has changed the way students access and interact with learning resources and other actors in the university, as well as opening new opportunities in learning and teaching processes. In parallel, new movements such as OER (Open Educational Resources) and the MOOC (Massive Open Online Courses) phenomenon generate changes in the area of higher education that require careful consideration for their integration into our educational model.

In this scenario, one of the challenges is how to analyze all these changes from a perspective wide enough to understand the relationships among them as well as the implications in the student and university interaction. This is why the eLearn Center aims to provide both the framework and the ground for implementing these changes in a controlled environment, measuring their impact, collecting the necessary data and evaluating their scalability in order to convert every teaching experience into a good practice, facilitate rapid adoption and dissemination. In doing so, the eLearn Center ensures the continuous contribution of the UOC to e-learning research and innovation.

Consequently, all the eLearn Center activities follow a continuous cycle of analysis, design, implementation and evaluation, focusing not only on the student but also on the whole teaching/learning process, which is generated in the interaction (in a wide sense) between users (mainly students, teachers, tutors and mentors), services and resources. It should be possible to capture this interaction, analyze it to detect strong/weak points and, to propose solutions that may be swiftly implemented and evaluated. In this way the acquired knowledge is reintroduced into the learning scenario, repeating this process as needed.

In this paper we describe how the eLearn Center addresses the problem of both supporting and analyzing UOC's educational model whilst providing teachers, researchers and practitioners with an experimental space where they can design and implement new educational practices. We provide also an example of the kind of analysis that can be carried out within this framework.

The UOC as a laboratory

The UOC, as a result of a bold and successful decision to bring its activity online from the moment it was founded in 1994, has become a laboratory for e-Learning, and expertise in applied research in eLearning can only follow from this. Such expertise is not based on the reputation of individual researchers' production only, but on the university's ability to provide an experimental ground that can attract researchers of the field. The role of the eLearn Center is to plan for the full realization of UOC as a e-learning laboratory potential by means of managing effectively innovation synergies, supporting big data analysis, and having an active role in the design of our educational model, internal faculty/staff training schemes and a strong, innovative and well-informed technological strategic plan that can match the UOC's first decision to become a virtual university. Therefore, the eLearn Center provides support to the following domains:

- UOC educational model: this area is in charge of the application of the educational model as well as the evaluation of its implementation. It ensures the constant update and evolution of the model with the definition of instruments, environments and the methodologies that comprise it. In order to do so it is informed by applied research results and educational innovation developments, at the same time as it incorporates the latest pedagogical and technological trends.
- Innovation: the mission of this area is to promote an innovative university, as opposed to a university which only does innovation. With this objective, using a combination of bottom-up and open innovation approaches, as well as transversal and strategic projects, a culture of innovation is created.
- Applied research: the objective of this area is to facilitate and encourage the use of the UOC as a laboratory by teachers and researchers taking advantage of the data available on learning and teaching processes with ICT. The priority here is to foster research about the activity of the UOC itself, and applied to the areas of knowledge that are taught in the institution.

The eLearn Centre's Virtual Laboratory offers instruments and a space for experimentation to the three areas above described with the aim to foster change. It is a virtual space with the infrastructure, instruments and tools necessary for designing, executing and analysing new scenarios in e-learning. The e-Learn Center Lab, thanks to a Cloud Computing concept, allows teachers, researchers and management staff in the UOC to experiment with new initiatives and pilot them in a flexible but yet controlled environment. It is also in this Lab where the overall strategy for the use of learning analytics in applied research on the UOC eLearning activity is designed.

Managing big data

As summarized by Siemens and Gasevic (2012), learning analytics is defined as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs. ... High expectations exist for learning analytics to provide new insights into educational practices and ways to improve teaching, learning, and decision-making." The importance of this new research field is demonstrated by the enormous attention that it has recently received, as well as the existence of several conferences and journals devoted to this topic.

One of the main concerns for any educational institution aiming to do learning analytics on its data is, precisely, how to capture and maintain such data in a simple datamart¹ that allows researchers to further analyze it according to the designed experiments. Obviously, the intrinsic nature of each institution determines which data is available for analysis purposes. If learning analytics is an instrument that the

¹ A datamart is the access layer of the data warehouse environment that is used to get data out to the users.

eLearn Center uses to better understand and improve UOC's educational model, we need to establish different levels of analysis, depending on the nature of each problem being addressed and the context, in this case online / distance higher education. In the case of a virtual learning environment such as UOC's Virtual Campus, the amount of available data is huge; at every moment there are tens of thousands of users interacting with thousands of services and resources, thus satisfying the three Vs of the "big data" definition: volume, variety and velocity. This generates a large (and sparse) cube of data that cannot be analyzed as a whole; we need to define different levels of analysis (Mor et al., 2007) according to the desired research goals, as follows:

- Session level: what does the learner do when s/he connects to the Virtual Campus? This level captures the way users navigate with particular goals in mind. For example, how users use the e-mail service or how they access to the proposed exercises. At this level, the short-term navigation behaviour is studied. In this case, a web mining analysis could be helpful to detect problems with the web interface, for automatic usability evaluation purposes (Ivory and Hearst, 2001).
- Semester level: what does the learner do during an academic semester? This level tries to join all the single user sessions in a continuous flow during a longer period of time, with a limit of an academic semester. This medium-term navigation behaviour will be useful to validate hypothesis about the relationships of user actions and her results, which are related to the way learning resources are organized. The main goal of this level is to determine the navigational patterns followed by users but at a higher scale than in the previous level. For example, it is interesting to study whether students connect every day or not, or whether they make an extensive usage of the virtual classroom forums during the weekend or not. All the information collected at this level could be used to feed an intelligent tutoring or adaptive hypermedia system; see (Brusilovsky, 2001) for instance.
- Degree level: what does the learner do from the time of enrolment until s/he finishes (or abandons) the degree? Finally, this level can be considered a long-term navigational behaviour analysis. In this case, the main interest is to analyze how students evolve from the beginning of a degree until they successfully finish it (or less successfully, they drop out). This includes the study of several stages in the student life-cycle: first contact and university access, first and following registrations, etc. Performing a data mining analysis at this level could help tutors and mentors to choose more carefully the subjects each student is enrolled to each semester, for instance. At this level it may be also interesting to discover inappropriate combinations of subjects that might lead to an excessive teaching burden.

Each one of these levels needs different data in order to answer the established research questions. In fact, the Virtual Campus is a rich scenario for experiment design, as different research questions involving different analysis levels can be imagined. Depending on the available information (collected usage data, surveys, academic results, etc.) and the desired goals, different experiments can be designed. Following the same approach, the datamart can be designed as follows:

- At session level, each Virtual Campus service executed by a user generates one or more tuples according to the following abstract model:

$$(U \langle id \rangle, T \langle timestamp \rangle, S \langle id \rangle, R \langle id \rangle, X \langle id \rangle)$$

That is, user U (identified by <id>) in moment T applies service S to resource R with result X. If all available services at the Virtual Campus follow this approach, it would be possible to create a small datamart for each service (or group of services), thus simplifying further analysis. That is, the institutional learning environment could define a set of available services and store only all data regarding such services to a specific datamart. On the other hand, if there is an experiment involving the use of an intelligent tutoring system for a particular course, those data will be stored in a different datamart. With the support of the eLearn Center, each teacher, researcher or practitioner will need only to define the semantics for the set of services, their scope (that is the resources where the service can be applied to) and the possible results for a given experiment.

The eLearn Center laboratory will provide the place where to implement and deploy such experiments.

- At semester level, we can obtain additional information regarding user activity: her academic performance, her motivations, expectations and satisfaction (by means of one or more surveys) and a summary (i.e. aggregated) of her navigational behaviour using the aforementioned session level data. All these data also uses the same abstract model previously defined. In this level we are only interested in what is happening in the current semester, thus keeping a reasonably small datamart.
- At degree level, we maintain a huge datamart with all historical data from previous semesters, including also any other information relevant to user profiles (i.e. age, gender, previous academic background, special requirements, and so) in different databases.

Therefore, experiments are executed in the laboratory, feeding one or more session and semester level datamarts. At the end of each semester, once all the desired data has been gathered, the semester level is dumped on the degree level (or historical) datamart. Depending on the research questions, the researcher does not need all the available data but only those related to her level of analysis.

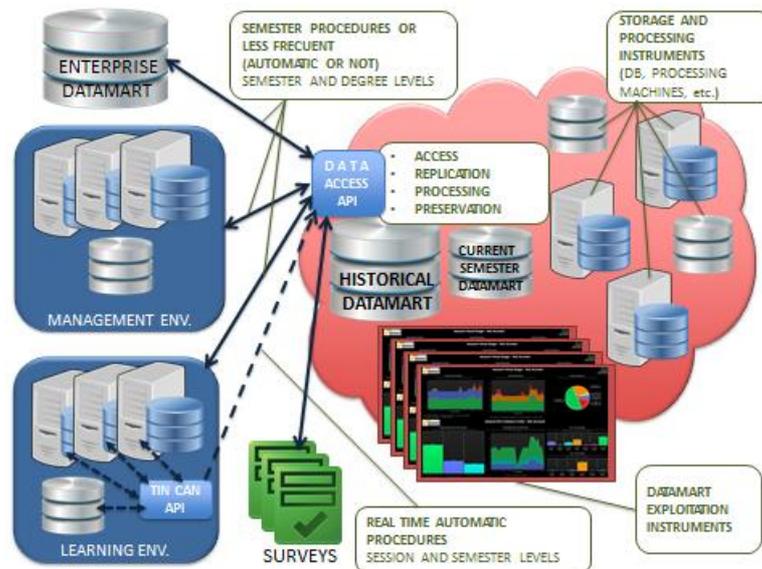


Fig1. The UOC's Learning Analytics infrastructure and procedures

Case study: analyzing dropout

University dropout is a major issue and should be seen as a failure of the higher education system to create an outcome (graduates) after having invested a significant amount of resources, normally publicly funded (OECD, 2012). However, the financial costs of dropout are only part of the total costs: Non-pecuniary (or affective) costs –more difficult to measure – are also important for dropout students (Johnes, 1990). It should be noted that university dropout is a multidimensional phenomenon that needs to be correctly defined before any thorough analysis of its causes can be carried out. One of the authors who has put great emphasis on creating a university dropout doctrine is Vincent Tinto (1975). Tinto stresses the importance of reaching a good definition of university dropout, which he sees as essential as detecting the causes of this dropout. In a more recent vision, Lee and Choi (2011) have reviewed research on online course dropout. The results of this analysis reveal that so far research has focused mainly on analyzing the causes of dropout on a course level.

Analyzing dropout is a perfect example of how the different data gathered and stored in one or more of the available datamarts can be combined into a single dataset for analysis purposes. Preliminary experiments (Grau-Valldosera and Minguillón, 2014) show that taking a break during the second semester is a clear sign risk of dropout. These experiments use the following data:

- From the degree (historical) datamart:
 - User profile: age, gender, work status, family status and so. This information is gathered during the first enrolment (the first academic semester) and stored in a database within this datamart.
 - Academic background and previous higher education experience.
 - Courses enrolled during her first semester and the academic performance for each course. This might include aggregated indicators summarizing the activity in each course, i.e. the number of times the user has interacted with a specific tool or resource.
 - Aggregated data from her navigational behaviour: number of connections to the Virtual Campus, average connection length, etc.
 - Any other aggregated data summarizing the use of other services available through the Virtual Campus, such as the number of accesses to the Digital Library, the interaction with her tutor, and so.
 - Finally, any other information that may be gathered through surveys during the first semester, including motivations to study the degree, expectations and so. Notice that this information may not be available for all users.

- From the semester (current) datamart:
 - The fact whether the user is enrolled or not in the second semester.

With all these data, researchers are able to build complex models trying to discover and explain underlying patterns and reasons about why students drop out. The eLearn Center acts as a promoter, curator, facilitator, data custodian and disseminator, providing a holistic framework for both research and innovation in the field of online education.

Conclusions

Although both traditional brick-and-mortar universities and online / distance ones have been using virtual learning environments since some years ago, it has only been recently that such institutions have seriously considered the opportunities that the analysis of such huge amounts of data brings. In this respect the large volume of learners and the intensive use of educational technologies made by the virtual campus users place the UOC in a unique position to carry out evidence-based applied research. For this it is necessary that UOC researchers have at their disposal the needed instruments and well defined protocols for to gather and exploit data for the research and improvement of learning and teaching processes. This kind of applied research is key for one of the main missions of the eLearn Center, that of guaranteeing the quality of the UOC educational model by evaluating it and improving it continuously.

In UOC, the eLearn Center acts as a bridge between applied research on online education and technological development. "Pedagogy should not be subject to technology, this statement has been a great matter of concern and used as a sign of quality in e-Learning. Likewise, technology cannot be limited by pedagogical principles based on face-to-face teaching. The way technology has become part of our lives today influences how we communicate and interact with technology opening up new possibilities that may not have been possible in a f2f educational context. The true potential of e-Learning in the context of the UOC is realized when balance is achieved and both pedagogy and technological development walk hand in hand. Innovation is what makes this possible.

References

1. Brusilovsky, P. (2001). Adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 11(1-2):87–110.
2. Grau-Valldosera, J.; Minguillón, J. (2014). Rethinking dropout in online higher education: The case of the Universitat Oberta de Catalunya. In *The International Review of Research in Open and Distance Learning*, 15(1).
3. Ivory, M.; Hearst, M. (2001). The state of the art in automated usability evaluation of user interfaces. *ACM Computing Surveys*, 33(4):173–197.
4. Johnes, J. (1990). Determinants of student wastage in higher education. *Studies in Higher Education*, 15(1), 87.
5. Lee, Y.; Choi, J. (2011). A review of online course dropout research: implications for practice and future research. *Educational Technology Research and Development*, 59(5), 593-618.
6. Mor, E.; Minguillón, J.; Garreta-Domingo, M.; Lewis, S. (2007). A Three-Level Approach for Analyzing User Behavior in Ongoing Relationships. In *Proceedings of the 12th International Conference, HCI International 2007*, Beijing, China, July 22-27.
7. OECD (2012). How many young people graduate from tertiary education? In *Education at a glance 2012: Highlights*. OECD Publishing.
8. Sangrà, A. (2002). A new learning model for the information and knowledge society: The case of the Universitat Oberta de Catalunya (UOC), Spain. *The International Review of Research in Open and Distance Learning*, 2(2).
9. Siemens, G.; Gasevic, D. (2012). Guest Editorial - Learning and Knowledge Analytics. *Educational Technology & Society*, 15(3), 1–2.
10. Tinto, V. (1975). Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89-125.