Big Data & eLearning: A Binomial to the Future of the Knowledge Society

Vidal Alonso, Olga Arranz

Universidad Pontificia de Salamanca

Abstract — There is no doubt that in what refers to the educational area, technology is producing a series of changes that will greatly affect our near future. The increase of students experiences in the new educational systems in distance learning makes possible to have information related to the students 'activities and how these can be dealt with automatic procedures. The implementation of these analytical methods is possible through the use of powerful new technologies such as Data Mining or Big Data. Relevant information is obtained of the use made by the students of the technological tools in a Learning Management System, thus, allowing us to infer a pattern of behavior of the students, to be used in the future.

I. Introduction

As we all know, technology is having a great impact on people´s lives. If something has marked the progress of advanced societies, over the past few decades, this has been the remorseless development and massive use of technological tools to manage all kinds of tasks [1]. There is no doubt that in what refers to the educational area, technology is producing a series of changes that will greatly affect our near future. The recent emergence of MOOCS (Massively Open Online Courses) is just a sample of the new expectations that are offered to university students.

Another sample is given by the new channels of communication, which are represented by the social networks, which are beginning to be integrated into the educational system by means of lessons via video conferencing or participation in lively debates online. The exchange of information in Facebook or via messages on Twitter, allows a provision of information written according to the new standards [2]. To deal with these circumstances, teachers need to understand the new media available to them and use them creatively.

These educational developments provide the system with a large amount of data coming from the students' activity. The increase of student's experiences in the new educational systems in distance learning makes possible to have information related to the activities of the students and how these can be dealt with automatic procedures.

This trend leads to a role change in the behavior of the different educational agents, who both, teachers and pupils, must conform to the new methods and change their traditional ways of teaching [3]. Academic institutions can´t stay out of this phenomenon and are required to modify its structures and their information systems to meet the student's needs in order to have access to their academic offer.

Is there anything we can do with the vast amount of data provided by students to improve the educational system? Until relatively a short time ago, storage techniques did not permit an exhaustive analysis of the information present in the Learning Management Systems (LMS). Nowadays there are more and more new analytical methods that allow us to deal with the study of these data and infer trends of the use that students make with respect to the tools available in platforms.

In addition, it combines the implementation of these analytical methods with techniques, such as Big Data, that will enable the access of a new system which will provide new information about the students that we have in our classrooms [4]. The implementation of these analytical methods is possible through the use of powerful new technologies such as Data Mining or Big Data that enable the processing of large amounts of information by searching and finding out new knowledge that is present in the data [5]. Big data allow for very exciting changes in the educational field that will revolutionize the way students learn and teachers teach [6].

This is the context where this paper is framed. The use of the information provided by the new analytical learning techniques will provide assistance to the teachers who use the Learning Management System. To achieve this, we will work with a sample of data present in a LMS and we will see the use that the students have made of the available tools.

With this study, relevant information is obtained of the use made by the students of the technological tools which will allow us to infer a pattern of behavior of the students. This pattern may be used by teachers to achieve, through the application of technological teaching strategies, a greater motivation and productivity of students so that they are continually active in their learning process.

II. Big Data Inside Education

A proof of the importance that is reaching the phenomenon of Big Data is its implementation in educational institutions which are beginning to exploit and understand the benefits that it offers them. In the present time, just enterprises and organizations are the ones which have been analyzing these enormous sets of data to better understand their customers by trying to predict market trends, "the educational world is beginning to integrate the data sets available to improve the learning process of the students" [7].

At the beginning, the companies granted almost no value to the data collected in their transactions. When the Big Data Era began, institutions and business organizations became aware of the high potential remaining in the stored data in their files. This fact changes the trend toward the collection and data storage process, making a greater effort to maintain and structure their data repositories, those who are usually disorganized, contain unnecessary details and many times the knowledge locked in them is incomplete, being necessary a purification of them to avoid the generation of uncertainty [8].

The processing of the large amount of existing data in the field of education has been made possible thanks to the development of new Information and Communications Technologies (ICT). This development has led to diverse educational institutions to carry out an

analysis of existing data from the interactions of its students, and in this way, draw conclusions that will improve the working environment, generating new educational organizations structures, and what is more important, new learning processes.

In this sense, the NMC Horizon Report, a reference to the global level of the emerging technological trends in education, provides, in its last report of 2014, that "the data analysis shall be adopted, in a meaningful way, in a period of between two and three years, and, in fact, it is already used in some American universities." [9].

This adoption will be supported by the rapid deployment of the virtual learning environments and the MOOC (Massive Open Online Course), where students perform online tasks leaving a significant trail of data on the web. The collection and analysis of the collected data from transactions, that have made the students when interacting with the system, will be used to adapt the content to the students´ needs and thus, to act in the improvement of the education system.

The importance of the impact that Big Data is taking in the education sector is beginning to be reflected in the expectation aroused in a large percentage of teachers and researchers who have placed their hopes in that the analysis provide relevant data and what the use of these data would mean for the educational area.

Teachers must observe the behavior patterns generated and reduce the risk of students who give up, through a more personalized learning process. Therefore, the educational analytical process itself will detect new problems, generating possible corrections to improve the teaching-learning process, or even questioning the effectiveness of the teaching programs that are taught in the educational organization.

On the other hand, students also benefit because, thanks to the analysis of these data, teachers can adapt the learning environments to their needs. This environment adaptation will depend on the creativity of the teacher, who will interpret the patterns from each student and will choose to provide creative solutions that will help the student to learn the skills required.

Higher education has traditionally been inefficient in the use of data, often operating with substantial delays in analyzing readily evident data and feedback. Organizational processes often fail to utilize large amounts of data on effective learning practices, student profiles as well as providing interventions [10].

To analyze this immense amount of information two treatments or processes are beginning to be used increasingly, known as Data Mining and Big Data. Data Mining is also known as KDD process (Knowledge Discovery Databases), which can be described as a process, which allows us to discover hidden information in large volumes of data. In the course of the process it works with data subsets, looking for similar patterns of behavior or predictive models that can be inferred from the processed data. In the educational area it is used in a way that learning processes could incorporate new and relevant knowledge that enables improvements in such processes [11].

Analytics in education must be transformative, altering the existing teaching, learning, and assessment processes, the academic work, and administration tasks. Analytics provides a new model for university leaders to improve teaching and learning processes and will serve as a foundation for changes. But using analytics requires careful thinking about what we need to know [12]. In the same way, academic analytics has the potential to create actionable intelligence to improve teaching, learning, and student success to predict which students are in academic difficulty as well as focusing on specific learning needs [13].

While its use began with economic purposes, their multiple possibilities have allowed us to extend its use to the field of education. The main methods used and their key applications are [14]:

- *Prediction*: Develops a model to infer some aspects of the data. It

is used to emulate the behavior of students in the premises of their previous activities and to predict the possible outcomes.

- *Clustering*: Looking for classifying data into groups with the same characteristics providing information of common patterns for students who are in the same group.
- *Relationship Mining*: Finds out relationships among variables. It allows discovering associations of activities that can induce a sequencing of the same nature. It also highlights the most effective pedagogical strategies in the learning process.
- *Visualization:* It allows discovering trends in the use of educational platforms that are outside of the average of students, known as data noise.

However, with the eruption of MOOCS the online information storage is growing in such a way that the processes for managing this information is becoming insufficient, causing a serious problem due to not being able to exploit the data with the necessary guarantees [15]. In order to be able to process such information is necessary to have new methods, being Big Data the last to be applied to the learning area.

This method allows in the present time that organizations can capture and analyze any data, regardless of what type, how much, or how fast it is moving, and makes more informed decisions based on that information. In education, big data allows to understanding how students move through a learning trajectory. This includes gaining insight of how the student accesses the learning activities or measuring optimal practice learning periods [16].

We are very well aware of the fact that there is still a lot to learn about how to work with big data, just like everyone else. But one thing we know for sure is that the traditional ways of working with data will not lead to success in big data analytics [17]. The variety of information sources, the volume of information, latency of processing, even the basic business models are often all different in the big data space. Someone who recommends using the same old tools under these new circumstances is someone who is outside of the data analysis.

III. Big Data and Virtual Learning Environments

According to Castaneda, since the first Virtual Learning Environment was created in 1995, until today, they have made many mixed environments of telematic known tools like Virtual Learning Environments or Virtual Learning Environments (VLE) giving support different teaching and learning modalities available today. The education provided is configured using VLE, and takes shape through the so-called virtual campus or Learning Management System [18].

A lot of LMS are currently being used by companies, schools and universities to assist in their formation processes. You can even say that they have become the essential tool to perform a teaching model eLearning.

The use of these LMS in education eLearning involves generation of a large and complex set of data. These data obtained from the use by many users of the different technological tools in a virtual learning platform can be drawn great benefits to improve eLearning education.

How can one profit in the learning context from Big Data? Everyone interested in eLearning education wants to find the answer to this question. Ambrose pointed to the company in collaboration with IBM Skillsoft how big data can help create a learning experience more personalized and adaptive based on real information about each student. [19]

So, the study on the use of personal data can be applied perfectly to eLearning. It offers us the opportunity to learn more about our students and their behavior patterns in a way not known before. Moreover, we can use this knowledge to develop eLearning courses really geared

to the needs of our students through scenarios that meet their real situations.

Big Data offers us the opportunity to provide students with more efficient courses and more effective on-line learning modules which are attractive and informative. The reasons why large amounts of data can revolutionize the industry of eLearning are [20]:

- **It allows eLearning professionals design more customized eLearning courses.** If you give eLearning professionals the opportunity to learn what works best for their students, in terms of content and delivery, this will allow create more personalized and attractive eLearning courses, thus providing high quality and meaningful learning experience.
- **It provides counseling on effective online strategies.** ELearning big data can give us visions of which eLearning strategies work and which do not.
- **It allows the monitoring of students´ patterns.** With large data eLearning, educators gain the ability to track the students throughout the learning process. This helps them to find out patterns that not only will allow them to learn more about the behavior of each pupil, but also of the group of students as a whole.
- **It enables the possibility to expand our understanding of the process of virtual learning.** It is essential that eLearning professionals get to know how students learn and acquire knowledge. Big Data gives us the opportunity to gain a deeper understanding of the process of eLearning and how students are responding to the eLearning courses. This information can be used to design new learning methods.

Big Data provides teachers with highly relevant information, but we must not forget that it also brings benefits to students. For instance, if one of the benefits is that educators are able to produce better teaching materials to meet their learning patterns, the student will benefit from it as well, that is, if a student is presented with the information in a meaningful way he or she is going to be more motivated in their learning process. [21]

Students and participants in eLearning courses have much to gain from the benefits that the information from Big Data provides us. Next, a case study is presented based on the large amount of data collected using a virtual learning environment, where it is shown how eLearning and Big Data form a binomial that must be considered in the future to improve the knowledge society.

IV. A Case Study: Using Big Data in a Learning Management System

In this practical study it has been taken into consideration the need to evaluate the large amount of data generated from the use of the technological tools used in the teaching / learning environments interaction. These environments of interaction are related to the combined method of learning Known as "Blended Learning", which Integrates, in a balanced manner, the virtual classroom learning with the learning proposals.

The purpose of this analytical study is to provide information that will allow us to improve the outlook of teachers and students in order to optimize our design in eLearning courses. In addition, the study also aims at guiding the use of the most appropriate virtual educational tools, in order to develop innovative educational strategies according to the patterns obtained from the behavior of student learning.

The study takes, as its starting point, a sample of educational data, educational dataset, from a university that is accessed by the students through a virtual learning environment. The study analyzes the use that the students do of the different technological tools available when they

are accessing to the subjects enrolled.

In this study, the educational dataset provides the total number of accesses to the tools that has been collected in the data processing center of the university center. The total number of accesses to each one of the tools is shown in the Fig. 1:
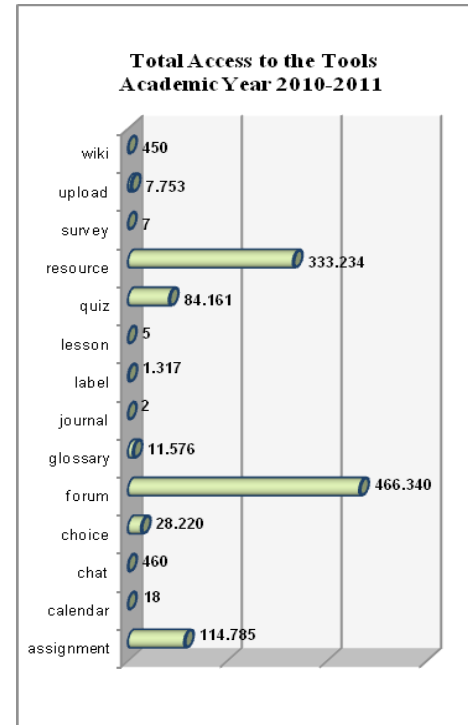

Fig. 1. Total Access to the Tools

As can be seen in Fig. 1 there are 14 different types of technological tools that can be accessed. There are other 2 tools, Course y User, which are not relevant to our study since they are related to the number of users who have accessed as well as to the available virtual courses.

In order to begin this research, it is necessary first to carry out a classification/organization in four different categories (Storage, Collaboration, Communication and Assessment), taking into account the use of the tools and at the same time the types of tools that an VLE should contain. This classification can be observed in Table 1.

TABLE 1. CLASSIFICATION TOOLS GROUPS

Categories	Tools Groups
Storage	Label, Resource, Upload
Collaboration	Forum, Glossary, Wiki
Communication	Calendar, Chat, Journal
Assessment	Assignment, Choice, Lesson, Quiz, Survey

However, it is important to highlight the Collaboration tool group, since it allows us to carry out a collaborative teaching-learning process providing feedback so that we can optimize the learning process as well as an increase in student´s motivation.

Once the group classification tools have been obtained, we can carry out an analysis of the average number of accesses of enrolled students to the different tool groups, always having in mind the final aim of the use of each tool group.

Taking into account the number of enrolled students and the number of total accesses to the different tool groups, we have obtained the

average number of accesses per each student and the type of tool used, getting, in this way the results shown in Fig. 2.
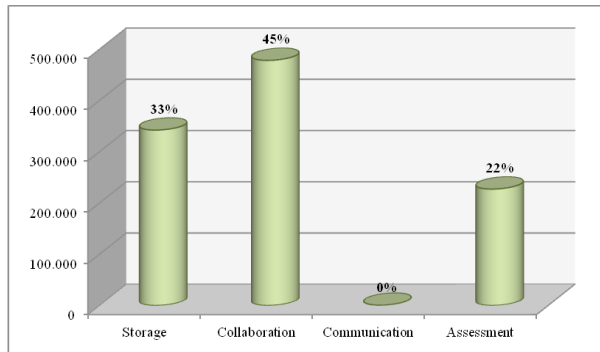

Fig. 2. Students Average Access by Tools Group

From the values shown in Fig. 2, it can be observed that the access to the Communication Tools represents 0% since the Calendar, Chat and Journal tools, which are within the Communication tools, are scarcely used.

On the contrary, Fig. 2 shows that the most used tools are the Collaboration 45% and Storage, 33%, whereas the rest 22% out of the total sample relates to the Assessment tools.

These percentages are aligned with accesses represented in Fig. 1 where it can be observed how the technology tools such as Forum, Resource y Assignment are the ones with a higher volume of accesses. Since they fall into three different categories, the percentage obtained doesn't show any significant differences.

It was also considered of high interest the need to perform a statistical analysis of the dependence or independence between different pairs of tool groups.

The results of this evaluation are shown in Table 2 where the results of asymptotic significance level and Chi-squared statistic obtained from the study of the corresponding media access for student and group sampling tools are presented.

Noted that in the table are represented only half of the values because the relationships between the types of tools are symmetrical blocks.

TABLE 2. REPRESENTATION OF THE VALUES AND SIGNIFICANCE OF THE ASYMPTOTIC CHI-SQUARE BETWEEN PAIRS OF TOOLS, FOR THEIR DEPENDENCE

	STORAGE	COLLABORATION		ASSESSMENT	
		SIG. ASYMPTOTIC	CHI-SQUARE	SIG. ASYMPTOTIC	CHI-SQUARE
STORAGE		0,000	82,96	0,000	148,01
COLLABORATION				0,000	35,08
ASSESSMENT					

In view of the results of the analysis about the asymptotic significance level, we can say that it will always exist dependencies between different tool groups represented. This statement is possible because the significance value in all of them is less than the reference value taken (0,005).

On the other hand, considering the data on chi-square analysis, it can be said that the greater dependence exists between the pair of Storage group with the tools of assessment, since the value of Chi-square is the greatest of all (148.01).

In contrast, less dependence exists between the group with the Collaboration tools and the Assessment ones, since the value of Chi-square is the smallest of all pairs of tools (35,08).

On the contrary, we can state that there is a linear relationship between accesses to the types of Storage tools with those accesses to Collaboration tools, thus, all who access the Storage tools also access the Collaboration tools.

So that, it is possible to extrapolate from data obtained that there is a very high dependence between the group of storage tools and the assessment one.

V. Conclusion

Given the huge amount of data available to the educational area, it is possible to proceed with its processing to obtain sufficient knowledge that allows them to improve their structures. The combination of different learning analytical techniques with the new paradigms of processing, such as Big Data, will enable relevant information to the educational authorities and teachers to change and to optimize the current methods.

These changes can be seen with fear on the part of teachers, who did not know how to deal with the new teaching methods from the pedagogical perspective. To help them, this paper shows a case study where the teacher gets information about which are the most used tools includes in the new learning environments.

From this information you can set new strategies of teaching-learning based on the student's experience. So that, looking for a greater participation by the students, the teacher may propose some tasks where they have to use the tools that are more favorite to him in front of other lesser-used tools.

The proposed activities, that involve the use of collaborative tools, makes the need to work the activity as a group, in line with the new educational circumstances that are supported, mainly, in the teaching-learning collaborative process. The use of these tools is highly satisfactory by part of the students which will result in a more active participation and an increase in the student's own motivation, driving to a learning improvement.

Furthermore, if these collaborative tools are combined with the storage and evaluation tools we shall be creating a teaching strategy that will not be rejected by the student, where he can develop all its intellectual capacity in an enjoyable and satisfactory way.

We have already seen how Big Data can influence eLearning, what benefits can bring and what the ways to use them are. But taking into account how fast this social trend progresses, we wonder if it is as good and beneficial as we are led to believe by the large companies, or, we might end up suffering the collateral consequences that no one has mentioned when describing the advantages of each product.

References

[1] O Arranz et al., *Surviving to the Education Online. Manual to Use Resources of Internet in the Classroom*. Salamanca, Ediciones Demiurgo, 2005.

[2] A. Pentland, "Una Sociedad dirigida por Datos*". Investigación y Ciencia*, pp. 46-51, Jan. 2014.

[3] C. Trevitt, E. Breman, and C. Stocks, "Assessment and Learning: Is it Time to Rethink Student´ Activities and Academic Roles?" *Revista de Investigación Educativa*, num. 30(2), pp. 253-270. 2012

[4] B. Tulasi, "Significance of Big Data and Analytics in Higher Education." *International Journal of Computer Applications*. 68, 14, pp. 21-23. April. 2013.

[5] J. Dean, *Big Data, Data Mining and Machine Learning*. Hoboken, New Jersey: John Wiley & Sons, 2014.

[6] M. V. Rijmenam. (2015, April 29). Four ways Big Data will revoluzione Education. *Datafloq*. [Online] Available: https://datafloq.com/read/big-data-will-revolutionize-learning/206

[7] O. Sánchez. (2014). Learning Analytics, el Big Data en Versión Educativa.

El Blog d'UPCnet. [Online]. Available: http://blog.upcnet.es/learning-analytics-el-big-data-en-version-educativa/

[8]   L. Joyanes, *Big Data: Análisis de Grandes Volúmenes de Datos en Organizaciones.* México: Mancorbo, 2014.

[9]   L. Johnson, S. Adams, V. Estrada, A. Freeman, *NMC Horizon Report: 2014 K-12 Edition.* Austin, Texas: The New Media Consortium, 2014.

[10] IBM, *Analytics: The Real World Use of Big Data. How Innovative Enterprises Extract Value from Uncertain Data. Executive Report,* IBM Institute for Business Value. 2013.

[11] J. Campo-Avila, R. Conejo, F. Triguero, R. Morales, "Mining Web-based Educational Systems to Predict Student Learning Achievements" *International Journal of Interactive Multimedia and Artificial Intelligence,* Vol. 3, No. 2, pp. 49-54. 2015.

[12] R. Ferguson. (2012). The State of Learning Analytics in 2012: A Review and Future Challenges. *Technical Report KMI-12-01,* [Online]. Available: http://kmi.open.ac.uk/publications/techreport/kmi-12-01.

[13] P. Campbell, P. Deblois and D. Oblinger , "Academic Analytics: A New Tool for a New Era." *EDUCAUSE Review,* Vol. 42, No. 4, pp. 40-57. 2007.

[14] R. Baker, *Data Mining for Education.* Pittsburg: Carnegie Mellon University. 2009.

[15] F. Jacobi et al.. "Towards a design model for interdisciplinary information systems curriculum development, as exemplified by big data analytics education." In *Proc. of the ECIS 2014, Twenty Second European Conference on Information Systems,* Tel Aviv, June 2014

[16] V. Kellen et al.." Applying Big Data in Higher Education: A Case Study." *The Data Insight & Social BI, Executive Report,* 13 (8). 2013.

[17] D. M. West. (2012). Big Data for Education: Data Mining, Data analytics, and Web Dashboards. *Governance Studies, the Brookings Institution.*[Online]Available: http://www.brookings.edu/research/papers/2012/09/04-education-technology-west

[18] L. Castañeda, "Entre Construir Entornos Virtuales de Enseñanza-Aprendizaje y Enriquecer Entornos Personalizados de Aprendizaje," presented at the EDUTEC2008 Conf., Santiago Compostela, 2008.

[19] J. Ambrose. (2014, Jan 21) Strategy, Corporate Development and Emerging Business, *Skillsoft,* [Online]Available: http://www.skillsoft.com/about/press_room/press_releases/january_21_14_ibm.asp

[20] Ch. Pappas. (2014, Jul 24). Big Data in eLearning: The Future of eLearning Industry, [Online] Available: http://elearningindustry.com/big-data-in-elearning-future-of-elearning-industry

[21] R. Scapin, (2015, Mar 29) Big Data in E-Learning: Looking To the Future, [Online] Available: http://freevideolectures.com/blog/2015/03/big-data-in-e-learning-looking-to-the-future/

Vidal Alonso was born in Luanco, Spain, in 1966. He received the Computer Science Degree in 1992 from the Polytechnic University of Madrid, and the Ph. D. degree, in 2004 from the Pontifical University of Salamanca, Spain. He was a Full Professor of Computer Science at the Pontifical University of Salamanca since 1994. He has occupied the position of Vice rector at his University for five years, until 2015. He also was the Director of the Computer Science School for nine years (2000-2009.) He works in data structures, knowledge discovery and data quality. Dr. Alonso is a member of ALI (Computer Science Spanish Association) and he won the Castilla y Leon Digital Award in 2007.

Olga Arranz was born in Burgos, Spain, in 1973. She received the Computer Science Degree in 2010 from the San Antonio Catholic University of Murcia, and the Ph. D. degree, in 2013 from the Pontifical University of Salamanca, Spain. She was a Full Professor of Education Department at the Pontifical University of Salamanca since 2002. She has occupied the position of Secretary Technical Director at his University for four years, until 2015. She works in elearning processes, knowledge discovery and educational technology.