Learning Analytics and Educational Data Mining: A Survey from 2005 to 2015

Chhavi Rana

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10683 Athens, Greece

ATINER’s conference paper proceedings series are circulated to promote dialogue among academic scholars. All papers of this series have been blind reviewed and accepted for presentation at one of ATINER’s annual conferences according to its acceptance policies (http://www.atiner.gr/acceptance).

© All rights reserved by authors.
Learning Analytics and Educational Data Mining:
A Survey from 2005 to 2015

ABSTRACT

Higher Education is at a point of unparalleled ambiguity and transformation with financial changes leading to increased focus on student focused model that emphasize on student engagement that leads them to better performance and employability [5,6,8]. The stakeholder in Indian Higher Education system faces stiff competition from International Universities and other organizations that are offering flexible education online. Learning analytics (LA) is the process of using this data to improve learning and teaching and refers to the measurement, collection, analysis and reporting of data about the progress of learners and the contexts in which learning takes place. Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in. In this paper, a comparative study is carried out using the output from projects implementing learning analytics around the world and there is an attempt to compare the strategies both LAK and EDM are using to contribute in this field.

Keywords: Critical thinking, Pedagogy, Learning analytics, Quality indicators, Group concept mapping.
Introduction

Data-driven decision making, popularized in the 1980s and 1990s, is growing into a more sophisticated concept recognized as big data that works closely with software approaches that are usually called as analytics. Big data and learning analytics for Higher education are in their formative years still their existence is already making noise and should not be overlooked. While Big Data and Learning Analytics are not magic bullet for resolving all of the issues and decisions encountered by higher education administrators, yet they have the ability to hugely aid the entire process of smoothing the administrative and instructional functions. The purpose of this study is to examine the evolving world of big data and learning analytics in Indian higher education context. The first report in this area was presented by Campbell and Oblinger (2007) [12] that linked technological techniques to the administrative approaches to a larger scale: “Learning analytics marries larger data sets with statistical techniques and predictive modeling to improve decision making” with “the potential to improve teaching, learning, and student success”. In the last few years, learning analytics (LA) has been growing progressively and emerged as a separate area of research. According to Siemens (2011) [36] LA is “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.” This research area borrows its concept from an amalgamation of fields ranging from educational data mining, process mining, data processing, technology-enhanced learning, information retrieval, and visualisation. It is emerging as a multi-disciplinary research field claiming its own grounds. Currently, there are several organisations that are working with this topic in conferences, workshops, special journal issues. There are courses, summer institutes, and societies (SoLAR, 2014) explicitly dedicated to LA. The research community has grown over the year and the area of learning analytics has developed state of art in its definitions, frameworks, processes, and challenges [37-43]. The usage of learning analytics can be extremely useful to learners as well as educators. It can help learners to better plan their requirements and educators to reflect upon learning assessments and thereby by synchronising the learning processes [20]. Colleges and universities need to meet a number of challenges that impact their quality with the passage of time and use of technology is in forefront in the last decade. This fact is highlighted in a report [22] that states that the elite position that American higher education has enjoyed for decades is now being challenged. Similarly, UK JISC, a registered charity that champions the use of digital technologies in UK education and research, launched the Business Intelligence (BI) Programme in between 2011 and 2012 as part of its JISC strategy 2010-12 to help educational institutions develop and use their corporate and business systems efficiently and effectively [1].

Technologies can play a key role here as new digital, learning, and analytics tools now make it possible to replicate some features of high-impact activity inside classrooms, whether through the design of inquiry-based learning or through the ability to access and manipulate data, mount simulations, leverage “the crowd” for collaboration and social learning, or redesign when and how students can engage
course content [10]. The growth in adoption of education technologies such as learning management systems (LMS), student information systems, social media, and lecture capture has enabled access to large stores of data directly related to learning and teaching practice [10, 11]. Higher education has to rely on the political and economic sectors of the country to resolve this complex situation. However, there are issues that higher education can address in terms of expanding educational opportunity and thereby improving employment opportunity that go beyond opening doors for students. Higher education must strive to ensure that access means students can complete degrees. Online learning is part of the solution to this problem but student attrition in colleges and universities is at unacceptable rates and needs to be addressed as well. Data-driven decision making is already being used to help colleges identify and evaluate strategies that can improve retention. As data-driven decision making enters the big data and learning analytics era, these new approaches, while not silver bullets, may be part of the solution. Higher education administrators will do well by evaluating whether they can be used in their institutions and determining the role they can play [2].

This paper begins with an introduction providing context and definition surrounding LA. Next, the literature review examines previous attempts to capture models of LA implementation. While we draw on and find much value in the dimensions identified in the literature, it is nevertheless apparent the existing models often represent LA as a linear, stepwise process. In our study we examine the degree of complexity that is associated with LA projects. Here we build a rich picture of actual and planned implementations in the Australian, American and UK Universities. This study will try to offer a unique perspective on the key capabilities required for an enduring adoption of LA, one that ensures the flexible capacity necessary to meet foreseeable demands, and demonstrates value for institutions, educators, and students. On the basis of this review we identify future challenges in design and implementation of LA. Finally, we will conclude our results and provide some limitations and potential future research directions toward the application of the quality indicators in learning analytics.

Literature Review

Research on learning analytics has started in the international arena since early 2000s [12]. Furthermore, researchers in the community have been working in the area of Educational Data Mining a decade ago. Both theoretical and experimental results are reported in the literature. Some of the work in the last decade is summarized below:

Ong, Vincent Koon. "Business Intelligence and Big Data Analytics for Higher Education: Cases from UK Higher Education Institutions.” Information Engineering Express 2.1 (2016): 65-75. “In this paper, Case study of UK Higher Education Institute is carried out to study the strategic use and applications of big data in higher education and whether it would lead to higher educational quality and better student and staff experience. This paper presents eleven cases of big data analytics in the UK Higher Education institutions, followed by presenting one of the
projects to demonstrate the application of big data analytics on student engagement”.

Sclater, Niall, Alice Peasgood, and Joel Mullan. "Learning Analytics in Higher Education." (2016). “This report documents the emerging uses of learning analytics in the United States, Australia and the United Kingdom. Through a series of eleven case studies it presents an overview of the evidence currently available of the impact that analytics are having on teaching and learning and highlights some of the opportunities for the UK higher education sector. Given the current focus on teaching excellence in the higher education sector, it will be of interest to policy makers and institutional leaders alike”.

Arroway, Pam, et al. Learning Analytics in Higher Education. Research report. Louisville, CO: ECAR, March 2016. 2016 EDUCAUSE, 2015. “This report states that although learning analytics lags institutional analytics, internal and external pressures are clearly increasing for institutions to use analytics to assess student success and resource management. It also states that colleges and universities have an opportunity now to proactively establish processes, understanding, and the use of analytics and establish themselves as the owners of and drivers in the future of learning analytics”.

Milevski, Zoran, Elena Gelova, and Zoran Zdravev. "Big data for education data mining, data analytics and web dashboards." Yearbook 2014-Faculty of Computer Science 3.3 (2015): 39-46. “This paper suggests that development of computerized learning modules enables assessment of students in systematic, real-time ways. It also stresses on the fact that Data mining and data analytic software can provide immediate feedback to students and teachers about academic performance. Therefore, such an approach can analyze underlying patterns in order to predict student outcomes such as dropping out, needing extra help, or being capable of more demanding assignments. It can identify pedagogic approaches that seem most effective with particular students”.

Mattingly, Karen D., Margaret C. Rice, and Zane L. Berge. "Learning analytics as a tool for closing the assessment loop in higher education." Knowledge Management & ELearning: An International Journal (KM&EL) 4.3 (2012): 236-247. “This paper examines learning and academic analytics and its relevance to distance education in undergraduate and graduate programs as it impacts students and teaching faculty, and also academic institutions. The focus is to explore the measurement, collection, analysis, and reporting of data as predictors of student success and drivers of departmental process and program curriculum. Learning and academic analytics in higher education is used to predict student success by examining how and what students learn and how success is supported by academic programs and institutions. The paper examines what is being done to support students, whether or not it is effective, and if not why, and what educators can do. The paper also examines how these data can be used to create new metrics and inform a continuous cycle of improvement. It presents examples of working models from a sample of institutions of higher education: The Graduate School of Medicine at the University of Wollongong, the University of Michigan, Purdue University, and the University of Maryland, Baltimore
County. Finally, the paper identifies considerations and recommendations for using analytics and offer suggestions for future research”.

Picciano, Anthony G. “The Evolution of Big Data and Learning Analytics in American Higher Education.” Journal of Asynchronous Learning Networks 16.3 (2012): 9-20. “This work examines the evolving world of big data and analytics in American higher education. It also looked at the nature of these concepts, provide basic definitions, consider possible applications, and last but not least, identify concerns about their implementation and growth”.

In India, We are not able to trace any researcher that is working in particularly the domain of higher education. Related to the Big and analytics, following contributions are traced in the national level:

Chandarana, Parth, and M. Vijayalakshmi. "Big Data analytics frameworks," Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014 International Conference on. IEEE, 2014. “In this paper, the important characteristics, issues and challenges related to Big Data management have been explored. Various open source Big Data analytics frameworks that deal with Big Data analytics workloads have been discussed. Comparative study between the given frameworks and suitability of the same has been proposed”.

Katal, Avita, Mohammad Wazid, and R. H. Goudar. "Big data: issues, challenges, tools and good practices." Contemporary Computing (IC3), 2013 Sixth International Conference on. IEEE, 2013. “This paper introduces the Big Data technology along with its importance in the modern world and existing projects which are effective and important in changing the concept of science into big science and society too. The various challenges and issues in adapting and accepting Big Data technology, its tools (Hadoop) are also discussed in detail along with the problems Hadoop is facing. The paper concludes with the Good Big data practices to be followed”.

Srinivasa, Srinath, and V. Bhatnagar. "Big data analytics." Proceedings of the First International Conference on Big Data Analytics BDA. 2012. “This volume contains the papers presented at BDA 2014: The Third International Conference on Big Data Analytics, held during December 20–23, 2014 in New Delhi. There were 35 submissions. Each submission received an average of 1.7 reviews from which the committee decided to accept 17 papers. BDA 2014 was the third in this conference series that began in 2012. The aim of this conference is to encourage dissemination of results from research and practice involving very large datasets”.
<table>
<thead>
<tr>
<th>Country</th>
<th>Institution</th>
<th>Project</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UK</strong></td>
<td>University of Central Lancashire</td>
<td>Student performance, retention and progression</td>
<td>Track student progress in course; intervention</td>
</tr>
<tr>
<td></td>
<td>University of Bolton</td>
<td>Budget and workload planning</td>
<td>Data management</td>
</tr>
<tr>
<td></td>
<td>University of East London</td>
<td>Student lifecycle and performance benchmarking</td>
<td>Track student progress in course; intervention</td>
</tr>
<tr>
<td></td>
<td>University of Sheffield</td>
<td>Student admission and progression</td>
<td>Student support and intervention</td>
</tr>
<tr>
<td></td>
<td>University of Durham</td>
<td>Performance benchmarking</td>
<td>Track student progress in course; intervention</td>
</tr>
<tr>
<td></td>
<td>University of Glasgow</td>
<td>Research interests and output</td>
<td>Enhance knowledge building</td>
</tr>
<tr>
<td></td>
<td>University of Manchester</td>
<td>Facilities and utilities optimisation</td>
<td>Student support and intervention</td>
</tr>
<tr>
<td></td>
<td>Liverpool University</td>
<td>Performance benchmarking</td>
<td>Track student progress in course; intervention</td>
</tr>
<tr>
<td></td>
<td>Open University</td>
<td>Student engagement, retention and progression</td>
<td>Track student progress in course; intervention</td>
</tr>
<tr>
<td></td>
<td>University of Bedfordshire</td>
<td>Student engagement, retention and progression</td>
<td>Track performance and predict student success</td>
</tr>
<tr>
<td></td>
<td>University of Huddersfield</td>
<td>Research performance</td>
<td>Track student progress in course; intervention</td>
</tr>
<tr>
<td><strong>AUSTRALIA</strong></td>
<td>University of Wollongong (UOW)-</td>
<td>The Social Networks Adapting Pedagogical Practice (SNAPP)</td>
<td>Student support and intervention</td>
</tr>
<tr>
<td></td>
<td>Edith Cowan University (ECU)-</td>
<td>Connect for Success (C4S)</td>
<td>Data management</td>
</tr>
<tr>
<td></td>
<td>University of New England (UNE)-</td>
<td>Automated Wellness Engine (AWE)</td>
<td>Enhance knowledge building</td>
</tr>
<tr>
<td></td>
<td>Open University</td>
<td>Personalised</td>
<td>Track</td>
</tr>
<tr>
<td>Australia (OUA) - Adaptive Study Success (PASS)</td>
<td>performance and predict student success</td>
<td>Data management</td>
<td></td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>----------------------------------------</td>
<td>----------------</td>
<td></td>
</tr>
<tr>
<td>University of Central Florida</td>
<td>EIS (Executive Information System)</td>
<td>Track student progress in course; intervention</td>
<td></td>
</tr>
<tr>
<td>Rio Salado Community College</td>
<td>PACE (Progress and Course Engagement)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northern Arizona University</td>
<td>GPS (Grade Performance System)</td>
<td>Student alerts for academic</td>
<td></td>
</tr>
<tr>
<td>Purdue University</td>
<td>Course Signals System</td>
<td>Student alerts for academic issues; intervention</td>
<td></td>
</tr>
<tr>
<td>Ball State University</td>
<td>Visualizing Collaborative Knowledge Work</td>
<td>Enhance knowledge building</td>
<td></td>
</tr>
<tr>
<td>University of Michigan</td>
<td>E2Coach</td>
<td>Student support and intervention</td>
<td></td>
</tr>
<tr>
<td>University of Maryland Baltimore County (UMBC)</td>
<td>Blackboard LCMS</td>
<td>Track performance and predict student success</td>
<td></td>
</tr>
<tr>
<td>Graduate School of Medicine, University of Wollongong</td>
<td>BIRT (Business Intelligence and Reporting Tools)</td>
<td>Reveal continuity of care issues</td>
<td></td>
</tr>
</tbody>
</table>

**Challenges and Future Directions**

Big data analytics can present valuable and exceptional insights about student engagement, student retention and progression issues. They can also be highly useful for assisting decision and making policy regarding research investment and impact analysis as well in optimization of university’s infrastructure and many other higher education related issues. The research in the area of big data and learning analytics for higher education can help in the building up new domains of higher education research that exists in other countries such as the Higher Education Information Database for Institutions (HEIDI) data, the Research Excellence Framework (REF) data, graduate employability data, National Student Survey (NSS) data for creating excellence in Higher education. Long and Siemens [44] stated a large number of benefits of using learning analytics for higher education. On one hand these benefits are focused on an administrative level, such as effective resource utilization through improved decision making, and thereby
increasing organizational productivity. On the other hand, LA can also aid faculty identify at-risk learners and thereby help in providing interventions at the right time, transform pedagogical approaches as needed, as well as assist students in identifying their own learning curve and requirements. Big data and LA thus can empower Higher education Institutes to reach greater heights in achieving excellence. The wave of Big data analytics in the higher education sector has been evolving through the concept of learning analytics (LA). This research area has been proclaimed as a revolutionary change for education sector as the results of LA implementations will deal with core education challenges. These comprises of the concerns regarding student performance monitoring, demonstration of learning and teaching quality, and developing models of personalised and adaptive learning. It can also be useful for performing a variety of administrative functions such as recruitment and admissions processing, financial planning as well [2, 12].

A survey done by NASSCOM stated only 25% of Indian IT graduate are employable and another report by Aspiring Minds stated similar claims for all graduate students [23]. The applications of Big data Analytics on teaching and learning can be highly helpful in this regard. Learning analytics is receiving increased attention, in part because it offers to assist educational institutions in increasing student retention, improving student success, and easing the burden of accountability [12]. Learning analytics is one of the six emerging technologies that are likely “to enter mainstream use” over the next five years as reported by a survey [18]. Hence, there is pressing need to research the effects of Big data analytics in Indian higher education context and how this opportunity to improve the performance of students of quality of education can be availed.

There are a number of key challenges in LA development and implementation. First of all, stakeholders such as students, teachers, administrators play a major influence in the success of LA development and implementation. There could be mismatched expectations among students, academic staff, and administrators, which could cause conflicts of interest and priorities in LA implementation in a university setup. A clear administrative structure must be put in place, followed by regular and continuous communication among key stakeholders for the successful implementation of LA. It was found that the BI solution may also change the strategic direction due to challenges in data usage and technological limitations. Other challenging issues include Data accessibility, data ownership, data quality and timeliness in LA development and implementation. Secondly, the availability and capability of technological solutions would determine the extent of advancement in LA, such as intelligent tools and techniques for proactive and intelligent information processing, interactive and dynamic digital visualization. Lastly, the development and implementation process must be user-centered from the beginning to the end. This will motivate and encourage students, academic staff and administrators to see the value and use the LA system for better management and strategic planning in Higher education domain.
Conclusions

Data is often stated as one of the key trends driving change in the 21st century [4]. Big data refers to the large volume of data – both structured and unstructured – that we create and deal with every day and is often depicted through the ‘three Vs’: high-volume, high-velocity and high-variety. Organizations use predictive analytics to draw insights from big data that give them the ability to offer each customer a tailored service according to their requirements that help them improve their customer relationship management which in turn convert tangible and intangible benefits [1]. Several reports [7, 8, 10] have stated that educational institutions today can respond to many external pressures in an effective and dynamic way by better understanding and henceforth better meeting the needs of the student body, as both key participant and funder. The understanding part in this whole process is of utmost important and accessible especially with the advent of big data.

In context of educational institutions, Every time a student interacts with their university – be that going to the library, logging into their virtual learning environment or submitting assessments online – they leave behind a digital footprint. Learning analytics is the process of using this data to improve learning and teaching and refers to the measurement, collection, analysis and reporting of data about the progress of learners and the contexts in which learning takes place. Using the increased availability of big datasets around learner activity and digital footprints left by student activity in learning environments, LA can provide extremely helpful insights in improving students’ engagement. Several reports document the emerging uses of learning analytics in the United States, Australia and the United Kingdom. For example, in a recent report of January 2016, From Bricks to Clicks, the Higher Education Commission in U.K concluded that analytics had “enormous potential to improve the student experience at university” and recommended that all institutions consider introducing an appropriate learning analytics system. This paper first discuss various types of learning analytics tool in academic and practitioner literature and proposed a conceptual framework that suit Indian Academia by building a synthesized set of definitions for analytics-related terms found in academia that are suited for Indian higher education. LA doesn’t just measure students’ progress, they can shape it and this paper has given a glimpse of the same.

References