INTERACTIVE LEARNING ENVIRONMENTS
2020, VOL. 28, NO. 2, 142–147
https://doi.org/10.1080/10494820.2019.1635395

Routledge
Taylor & Francis Group

EDITORIAL

Educational big data: extracting meaning from data for smart education

Nian-Shing Chen[a], Chengjiu Yin [b], Pedro Isaias[c] and Joseph Psotka [d]

[a]Department of Applied Foreign Languages, National Yunlin University of Science and Technology, Douliu, Taiwan; [b]Information Science and Technology Center, Kobe University, Kobe, Japan; [c]Institute for Teaching and Learning Innovation (ITaLI), The University of Queensland, Brisbane, Australia; [d]US Army Research Inst. (Ret.), Rockville, USA

Introduction

Contemporary education, at all levels, is unbounded by time and space and learning can occur both in physical and virtual environments and with the assistance of countless technologies and pedagogical instruments. Data are a constant in all learning transactions regardless of where they take place, resulting in an overwhelming volume of data sources and formats that are behind the concept of Educational Big Data (EBD). With the pervasiveness of technology and online learning, big data, in the context of education, has experienced exponential growth and includes data deriving from students' interaction with technology and their personal and academic profile (Ferguson, 2012). Furthermore, EBD is a central concern of educational institutions, as its value becomes increasingly visible and as a new body of evidence of its benefits becomes gradually published and disclosed among researchers and practitioners. Governments too will be able to use EBD to make educational improvements. As it is true for big data in other sectors, in education, one of the dominant conundrums of its existence is how to extract meaning from the data that is collected. While some agreement occurs in terms of the best instruments and techniques, governments, institutions and teachers are still left with the question of how to implement them and what questions to pose.

As the volume of data generated by education increases, more solutions for data management are required. Given the richness of the data that is collected in instructional settings, a growing number of educational institutions is using EBD for strategic planning and decision-making. EBD enables institutions to access data that is scattered in different sources; to respond more swiftly to the constant changes in the education sector; to make informed decisions based on data; to gain real-time insight into their students' behavior patterns and recommend solutions; to use predictive tools to enhance their students' learning results; and to support at-risk students EBD can be used by institutions to inform the development of educational policies by resorting to data-based decision-making (Picciano, 2012). By supporting the development of educational policy in data-based evidence, institutions can make decisions more objectively and with the support of the insight that the information extracted from EBD provides. They can more easily assess the different needs of their departments and courses and draft new policies accordingly.

Despite the fact that data is constantly being generated, and that it is available as a priceless resource, it exists firstly in an unstructured format. Hence, its significance depends on the capacity of reducing its multidimensional complexity into simpler relationships that can be used to improve the education system. Even though EBD is of significance for educational institutions, the education sector seems to be behind other sectors with concern to harnessing the benefits from the potential that analytics can represent. Educational institutions appear to remain far from having the capacity, at a practical, technical and financial level, to successfully master the collection, management and analysis of the EBD that is generated. Educational institutions need to restructure

CONTACT Joseph Psotka psotka@msn.com

their policy development processes and adopt innovative strategic planning (Macfadyen, Dawson, Pardo, & Gaševic, 2014). New policies and strategies will constitute a solid foundation for the adoption of the relevant instruments to harness EBD.

While the benefits of EBD are evident, it is important to acknowledge the range of challenges it poses, such as data security and privacy and the access to personal data (Wang, 2016). There are considerable costs involving the entire process of managing and analyzing big data, the institutional systems lack interoperability and the quality of the data that is collected is not guaranteed. Also, the meaningful generation of information from big data is a specialized skill that nonexperts find difficult to master (Daniel, 2015). The implementation challenges associated with EBD need to be carefully evaluated within each institution and measured against its benefits. The uniqueness of each institutions' capabilities prevents the outline of a universal solution and one-size-fits-all formulas. Instead, in an era led by data, institutions ought to examine the tools that are available to assist them in pursuing their missions.

Recently, Big Data have become mainstream in many research fields. The concept of Big Data "refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze" (Manyika et al., 2011, p. 1). Additionally, Big Data are often associated with key characteristics that go beyond the question of size, namely the 5 Vs: Volume, Velocity, Variety, Veracity and Value (Storey & Song, 2017). In recent years, EBD has become an important issue in the field of computer technology. With the emergence of online learning environments, such as OpenCourseWare (OCW) and Massive Open Online Courses (MOOCs), large volumes of data are being generated. Similarly, Learning Management Systems (LMSs) have caused a massive growth of the data that educational entities are required to manage. Since part of the students' learning occurs externally data is dispersed among various platforms that operate with different standards, providers and degrees of access (Ferguson, 2012). Furthermore, the data are produced in a variety of formats, such as image, video, text and audio. This abundance and diversity of data can be gathered and stored to be processed through analytic methods (Daniel, 2015).

Before the emergence of EBD courses and technologies, most research within classrooms mainly concentrated on the analysis of learning outcomes, with considerable limitations in terms of data and sample sizes. EBD focuses on extracting and analyzing meaningful information and as the data size is massive big data analytics tools must be used. Previous research has argued that the analyzing of EBD can facilitate the design of learning systems, educational materials and activities to improve educational effectiveness and optimize learning environments (Brajnik & Gabrielli, 2010; Greller & Drachsler, 2012; Law & Larusdottir, 2015; Sutcliffe & Hart, 2016). EBD analytics helps to discover novel and potentially useful information in large amounts of unstructured data. With the EBD analytics results, teachers and students can change their teaching/learning strategies. With the new strategies, new EBD will be generated and new EBD analytics results can continuously be provided to teachers and students (Hwang, Chu, & Yin, 2017).

Data are a valuable resource in education, with a panoply of applications across all educational levels, subjects and stakeholders. When data becomes overwhelming in terms of volume and complexity, innovative techniques are required to extract meaning and transform this data into valuable information. This special issue of the journal focuses on the value of EBD in the context of different education levels, from pre-school to higher education. The manuscripts included in this issue approach the potential of EBD for academic performance prediction, learning analytics implementation, and the identification and improvement of student behavioral patterns and performance. They are centred on students and emphasize the role that technology can play in improving their learning experience and performance.

Articles included in this special issue

In the manuscript "Pre-school children's behavioral patterns and performances in learning numerical operations with a situation-based interactive e-book", Li focuses on the use of technology as a

valuable tool to enhance children's mathematical literacy. The author reports on an experiment conducted in a kindergarten where one group of children worked with a situation-based interactive e-book system. The objective of the experiment was to evaluate the impact of the e-book on the children's learning performance of numerical operations. The results depicted a substantial improvement of the children's learning achievement within the group that used the e-book. Their attitude toward learning was also different from that of the children who did not use the e-book. The experimental group tried a greater number of strategies to complete the tasks that they were given, whereas their colleagues were more focused on peer communication. Macarini, in the manuscript "Towards the implementation of a countrywide K-12 learning analytics initiative in Uruguay", emphasizes the challenges that emerged during the development of a learning analytics study and tool that covered the entirety of the Uruguayan education system. Some of these main challenges subsume the heterogeneity of the education system, ethical and legal requirements and the integration and irregularity in databases. The author equally describes the design decisions and solutions that were used to address or minimize the problems that were faced and the three core experiments where they were applied.

In "Applying a fuzzy, multi-criteria decision-making method to the performance evaluation scores of industrial design courses", Juan Li addresses the limitations of conventional methods of performance evaluation of academic courses, by applying a fuzzy multi-criteria approach to decision-making. This method was intended to be a solution to respond to the complexity of the data in industrial design courses. The author conducted an experiment at a university with students and review experts using a fuzzy TOPSIS methodology (Technique for Order Preference by Similarity to the Ideal Solution) to evaluate the importance of seven evaluation criteria. The results indicated that the fuzzy, multi-criteria approach to decision-making can be employed to assess quantitative and qualitative data with a satisfactory level of objectivity and precision. Yang in the manuscript "Predicting students' academic performance by using educational big data and learning analytics: Evaluation of classification methods and learning logs" focuses on the capacity of learning analytics to identify students who are at risk and suggest an opportune intervention based on the results of their behavior analysis. The author posits that when deploying machine learning to train risk-identifying models, the factors that influence the performance of those models is overlooked. The results of the study that examined seven datasets within three universities reveal that the number of significant features, the number of categories of significant features, and Spearman correlation coefficient values are the factors influencing the predictive performance of classification methods.

Zhang in this manuscript titled "An individualized intervention approach to improving university students' learning performance and interactive behaviors in a blended learning environment" highlights the value of the data that derives from blended learning settings, which can be used to provide insight into the students' performance. The author designed a quasi-experiment to assess the effect of an individualized intervention applied to an experimental group of university students. The results showed that, in comparison with the control group, the experimental group demonstrated higher levels of motivation, learning attitude, self-efficacy and active learning behaviors. In addition, the results indicated that the experimental group had a better performance in terms of learning outcomes. Also within blended learning environments, Lai, in the manuscript "Effects of the group leadership promotion approach on students' higher order thinking awareness and online interactive behavioral patterns in a blended learning environment" conducts a study that measures the impact of a group leadership promotion approach that was applied to collaborative learning activities. The approach was assessed in a university by using an experimental group of students. When comparing the results from the control group, which learned with the traditional blended learning strategy, with the results from the experimental group, it becomes evident that the latter demonstrated an increase in their capacity for creativity, problem-solving, critical thinking and also higher performance.

Discussion

This special issue provides examples of EBD methodologies and tools to help researchers and educators to collect and analyze EBD. Most of it focuses on individual classrooms, but it has implications for institutions and more. We hope that broader issues will be addressed in the future. In order to offer readers a holistic overview of EBD, a summarization of the objectives, methodologies and potential research issues of EBD research are described as follows.

Objectives of EBD

1. Identifying or predicting students' learning status; recommending learning resources and activities; sharing and improving the learning experience.
2. Enabling educators to receive feedback, examining both the learning and the behavior of the learners, identifying the students who need support, determining which mistakes occur more often and improving the effectiveness of some activities.
3. Supporting course developers to evaluate the courses' structure and its impact on learning, assessing course materials, identifying the most valuable data mining methods according to different tasks and developing learning models.
4. Providing evidence to the administrators of educational institutions, helping them to organize resources, improving their offer of educational programs and assessing both teachers and curricula effectiveness.

Methodologies for EBD

With the growing popularity of EBD as research subjects, the goals have been placed on prediction, structure discovery, and relationship mining, and have used various methods to achieve those goals as depicted in Table 1 (Baker & Yacef, 2009; Yin & Hwang, 2018). In terms of prediction, a central research topic is predicting students' educational outcomes. In structure discovery, the emphasis is on finding structure, patterns and data points in a set of data without any ground truth or *a priori* idea of what should be found (Baker & Inventado, 2014). Relationship mining involves discovering relationships between variables in a dataset. These relationships are seen as rules of data for later use (Bousbia & Belamri, 2013).

Potential research issues

Despite the fact that there is a significant body of research in EBD, there are still numerous challenges and questions that remain to be addressed: how can EBD be built and used by integrating different kinds of data from online learning systems such as learning management systems and game-based learning systems? how to measure the adequateness and effectiveness of the EBD analytics results? how to promote the use of online learning systems for collecting EBD? how to protect personal privacy when collecting data from online learning systems, including privacy and security control policies?; how to employ EBD analytics results to support learning designs? how to integrate learning

Table 1. Goals and methods of EBD (Yin & Hwang, 2018).

Goals	Prediction	Structure discovery	Relationship mining
Methods	Classification	Clustering	Association rule mining
	Regression	Factor analysis	Correlation mining
	Latent knowledge estimation	Knowledge inference	Sequential pattern mining
		Network analysis	Causal data mining

theories and strategies with EBD? how to employ EBD analytics approaches in various application domains?

To provide further suggestions to researchers, a list of some potential research issues related to EBD are provided:

- Automatic assessment of student knowledge;
- comparing the behavioral patterns of the students with different personal factors, such as learning achievements, cognitive styles, learning styles or motives;
- data integration/cleansing methods and management tools for collecting meaningful EBD;
- data mining in social and collaborative learning;
- data mining with emerging pedagogical environments such as educational games, MOOCs;
- deriving representations of domain knowledge from data;
- detecting and addressing students' affective and emotional states;
- developing learning models or assessment models based on analytics results of EBD;
- evaluations and assessment of analytics results of EBD;
- evaluations of the efficacy of curriculum and interventions;
- generic frameworks, techniques, research methods, and approaches for EBD analytics;
- identifying students' behavioral patterns;
- identifying learning strategies from EBD;
- integrating data mining and educational theory;
- investigating the issue of personal privacy protection;
- Learning Design based on analytics results of EBD;
- multi-modal learning environments and sensor analysis;
- practices for the adaptation of analytics results of EBD to enhance teaching/learning environments;
- predictions and process mining from EBD;
- privacy and security management for open EBD;
- proposing new analytics' algorithms for learning environments or learning technologies;
- providing support for teachers and other stakeholders;
- theories and models in EBD analytics;
- defining and assessing the loop between education data research and educational outcomes;
- visualizations of learning activities with EBD.

Conclusion

The purpose of this special issue is to place an emphasis on the design, development and evaluation of using EBD. As EBD reaches the mainstream, it is crucial to examine design guidelines, best practices, and development methods that can assist educators to benefit from the potential of the data that is available and the techniques that can endow it with meaning. Moreover, once the design and implementation stages are completed, it is key to explore their effectiveness with methods that can provide insights into the actual value of using EBD. We anticipate that this special issue makes a clear contribution to those purposes and stimulate further research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Science Council, Taiwan under project numbers MOST-107-2511-H-224-007-MY3, and MOST-106-2511-S-224-005-MY3.

ORCID

Chengjiu Yin http://orcid.org/0000-0003-1492-5250
Joseph Psotka http://orcid.org/0000-0002-2359-3246

References

Baker, R. S., & Inventado, P. S. (2014). *Educational data mining and learning analytics. Learning analytics*. New York, NY: Springer. 61–75.

Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 3–17.

Bousbia, N., & Belamri, I. (2013). Which contribution does EDM provide to computer-based learning environments? In A. Peña-Ayala (Ed.), *Educational data mining applications and trends* (pp. 3–28). Dordrecht: Springer.

Brajnik, G., & Gabrielli, S. (2010). A review of online advertising effects on the user experience. *International Journal of Human–Computer Interaction*, *26*(10), 971–997.

Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, *46*(5), 904–920.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society*, *15*(3), 42–57.

Hwang, G.-J., Chu, H.-C. and Yin, C. (2017). Objectives, methodologies and research issues of learning analytics, Interactive Learning Environments, 25(2), 143-146.

Law, E. L., & Larusdottir, M. K. (2015). Whose experience do we care about? Analysis of the fitness of Scrum and Kanban to user experience. *International Journal of Human–Computer Interaction*, *31*(9), 584–602.

Macfadyen, L. P., Dawson, S., Pardo, A., & Gaševic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, *9*, 17–28.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). *Big data: The next frontier for innovation, competition and productivity*. McKinsey Global Institute.

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, *16*(3), 9–20.

Storey, V. C., & Song., I.-Y. (2017). Big data technologies and management: What conceptual modeling can do. *Data & Knowledge Engineering*, *108*, 50–67.

Sutcliffe, A., & Hart, J. (2016). Analyzing the role of interactivity in user experience. *International Journal of Human–Computer Interaction*, *33*(3), 229–240.

Wang, Y. (2016). Big opportunities and big concerns of big data in education. *TechTrends*, *60*(4), 381–384.

Yin, C., & Hwang, G. J. (2018). Roles and strategies of learning analytics in the e-publication era. *Knowledge Management & E-Learning*, *10*(4), 455–468.