The Role of Big Data in Improving E-Learning Transition

Karim Moharm¹, and Muhammed Eltahan²,³,⁴

¹Electrical Engineering, Alexandria University, Egypt
²Aerospace Engineering Department, Cairo University, Cairo, Egypt
³Now at Institute of Bio-Geosciences (IBG-3, Agrosphere), Forschungszentrum Jülich GmbH, 52425 Jülich, Germany
⁴Centre for High-Performance Scientific Computing, Geoverbund ABC/J, 52425 Jülich, Germany.

Email: karim.i.moharm@gmail.com, m.eltahan@fz-juelich.de

Abstract. Educational organizations are operating in a dynamic and highly competitive environment. Online learning, education, emerging technologies, and the number of learners are growing very fast and produce big data that contain meaningful information to be extracted and processed. The idea of applying the big data in e-learning supports developing and analyzing the students as well as the educational organization’s stakeholders. A big data framework architecture, methods, and tools in the e-learning platform are introduced. The paper studies briefly the opportunities of big data deployment in the educational sector and the value that the students and organizations gain. However, it also highlights the challenges and future research directions as the tool’s selection and value extraction from complex educational data sets.

1. Introduction

Educational organizations are operating in a very dynamic and highly competitive environment. The educational institutes have to have precise decisions to sustain its competitiveness. The traditional data warehouses in the educational field are going beyond the capabilities of traditional processing techniques for structured, semi-structured, and unstructured data sets and extracting the value beyond data sources as videos, audio, etc.

E-learning concept first aroused in 1990 and had no standard definition [1]. However, the American Society for Training and Development suggests that anything delivered using electronic technology for learning purposes [2,3].

The main obstacle for e-learning spreading is the active need for big educational data processing to have decisions based on analysis and not only intuitive or experience. Big data generated by learners, teachers, universities, institutions, etc. offer interesting applications in the educational field. Big data is described by the six V’s (volume, velocity, variety, veracity, value, and volatility). The volume refers to that big data processes large amounts of data that was impossible to process via traditional techniques. Data from thousands of learners who are enrolled in hundreds of courses in several organizations is a high-volume data issue. The velocity refers to the rate of obtaining the data is very high. For example, the data processing of a learner interacting with an educational platform that stores and analyses the streaming data gathered from the learner. The variety refers to the information that can be structured (learner profile data stored in rational databases), semi-structured (emails), or unstructured (free text, PDFs, videos, images, web-pages, RSS feeds, Facebook feeds, etc.). The veracity represents the truthfulness of the data and ensuring the data is reliable. The value indicates that the processed data is useful to end-users. The volatility shows that gathered data importance changes over time.

Institutes are gathering large amounts of data for each student to track performance and progress. Big data employment in the educational sector results in more academic achievements, less drop-out rate of students through continuous feedback and course adaptation, and effectiveness for the student and the teachers as
well. The effective performance analysis of the learner can also help in an effective recommendation for the student, along with the behavioral and historical data.

The remaining of this paper is organized as follows. Section 2 presents the big data sources in the educational sector and section 3 studies the background of big data tools and techniques. Section 4 discusses the proposed big data architecture layered in the e-learning, section 5 discusses big data opportunities and related work while section 6 proposes the challenges of implementing the big data in the education sector. Finally, the conclusion is drawn in section 7.

2. E-Learning Data Sources

Data in educational sectors is generated by several sources. Educational data sources include the academic data sets collected by the institutions, online public available data sets, and the communication data between different e-learning platforms. Other sources can be smartphones, web, media, audio, and blogs. Sensors such as wearable electronics, student attention sensors, and student electrical activity measurement sensors like NeuroSky [4] head tilt angle for learner engagement measurement are generating important data about the education process. Traditional data sets gathered by assessments and class drop rates are important data sources to be processed and analyzed. Online data sources through simulation and educational games [5] can contain important information about the learners. Educational environment interaction logs include MOODLE logs [6] which contain data about user ID, number of log-in trials, total time spent, user clicks, lessons progress, course information, and dated accessing of course materials [7]. Personal data sources include social media data such as university advertisements, online forums, student health status (disabled, diseases, medical reports, etc.), memory and concentration levels measurements, address, and financial situations. Historical data about the learner which involves Daily activities as workshop attendance, extracurricular activities as competitions participation, seminars participation, previous detailed grades, and previous schools’ history [8]. Classroom specific data such as the classroom cleaning status and bus routes are hiding big data that affect student learning [9].

3. Big Data Techniques & Technologies

3.1. Tools

Different tools are used for big data analysis, including data mining, analysis, and processing. The most used tools in the big data processing of e-learning are Apache Hadoop and Apache Spark. Apache Hadoop [10,11] is an open-source ecosystem for big data. It can be used as a cloud service. It contains several components such as Hadoop Distributed File System (HDFS), Yet Another Resource Negotiator (YARN), Pig and MapReduce. HDFS divides and replicates the data across multiple nodes. MapReduce [12] is responsible for mapping and reducing processes that are splitting tasks into chunks for hardware resources processing. Then, combining the outputs of the processed chunks of data. MapReduce programs can be programmed using Apache Pig and Pig Latin programming language [13]. Apache YARN [14] can replace the MapReduce process. Apache Mahout [15] is a scalable open-source machine learning library used for collaborative filtering, clustering, and classification. Mahout can be used with Hadoop. Apache Spark [16–18] is an open-source computing platform based on data analysis and computational memory capabilities for real-time and batch data processing. It is used by Yahoo, eBay, etc. It supports machine learning libraries as MLlib and visualization tools like GraphX.

3.2. Techniques

Several analytical methods are applied for the big data analysis and discovering value behind the data as the association rule mining, regression, classification (as decision trees, ANN, Bayesian networks) and statistical methods.

4. Design of Layered Big Data Framework in E-learning

Several analytical methods are applied for the big data analysis and discovering value behind the data as the association rule mining, regression, classification (as decision trees, ANN, Bayesian networks) and statistical methods. According to [19,20], the general framework for big data analytics in the e-learning can be through the following layers:
• Presentation layer: for visualization and statistical views. Output of the layer can be reports and graphs.
• Analysis layer: through MapReduce algorithms of the big data.
• Data pre-processing layer: data merging and data cleaning processes.
• Big data storage layer: through HDFS and HBase of heterogeneous data as learning platform data, card records, course videos, etc.

5. Opportunities
The output of big data analysis can be statistical analysis or visual presentation for the educational process and progress. Deploying the big data in the learning sector can enhance e-learning market and result in [9]:
• Dynamic tracking effectively the current students’ progress and performance per each subject using data sources such as personal data (ex. grades, hobbies, residence, activities, preferences), historical, behavior, and student interaction.
• Expecting the students who will fail and drop out of the course, especially in online courses [21].
• Preparation of personalized student development plan based on the student model, traits, behavioral analysis, and target of the learner whether to master the subject or learn an overview
• The digital textbook is emerging in a big data system. The digital textbook can be adapted based on the big data analytic of the student and learning style.
• Student interaction and engagement analysis with the class materials and provide modifications and enhancements, leading to student satisfaction [22].
• Marketing research about students and future market directions to attract more outstanding students.
• Improving student progress measurement and adaptive learning [23]. The traditional education system relies only on exam results to identify student progress. However, big data could generate progress metrics based on which questions are not correct and the meaning behind them, how long the student was thinking about the questions and selecting a wrong answer, and how many trials it took the student to get the correct answer.
• Providing relevant previous experiences from previous learners who faced the same issue that the current learner is struggling with. For example, big data analytic can recommend to the current learner different education lessons or insights that worked with previous learners with similar issues [24,25].
• Keeping the student in an energetic state through noticeable immediate actions implemented by big data analytic.
• Identification of each student learning depth of each concept.
• Courses enhancements through big data analytics of learners’ behaviors. If there are some very hard questions or materials that many students could not master, more explanation and modification of learning objects can solve the problem [22].
• Management and decision support for administration decisions as admissions, student profiles, alumni, strategic planning, industrial connections, campus management, and security monitoring [26].

5.1. Big Data Potential in E-Learning: Case Studies

5.1.1. Big Visualization, Dashboards, and Decision-Making Support: proposed Hadoop based system based on the student performance data warehouses. The system produces a visual presentation through graphs that represent the student academic progress for supporting top management decision making and statistical visualization dashboard for institution management.

5.1.2. Student Progress Tracking: suggested a random forest algorithm applied on big data for student grade forecasting based on Spark to monitor the student progress based on stored data in HBase based on different sources of data such as student personal data, academic history, absence rates, and previously detailed grades.

5.1.3. Educational Organization Quality Monitoring System: [27] designed a Spark-based big data solution for analysis of the institution’s big data such as daily instructional activities, student activities, academic results, etc. to monitor and evaluate the quality of the institution teaching and learning process and adjusting teaching strategies accordingly.
5.1.4. **Performance Prediction:** Big data enables the prediction of student performance and test grades based on the analysis of student information [28]. This can help student’s selection of best fit classes such as the application Degree Compass introduced by [29] which recommends the best class for the student to enroll.

5.1.5. **Choosing Future Career and Pathways:** Big data can support the future student decision for the future career. Zhenyu [30] proposed HDFS for data storage and MapReduce with Mahout to analyze exam results, student behavior data, and student life information for helping the student in Holland in the optimal selection of the future profession.

5.1.6. **Improved Personalized learning, Instruction, and Assessment:** Big data can adapt the e-learning learning objects. Big data analytics in the e-learning also reveals the power for providing personalized e-learning materials, student learning style identification, and adaptive recommending of learning objects [25,31]. Big data analytics can rapidly analyze learning difficulties and provide an opportunity to gradually improve and challenge the learner. [32] proposed a Spark-based system in C++ class to detect the hardest questions to students relying on the question expected difficulty and number of attempts to solve. [33] proposed a big data system for elective course recommendation using analyzing the grades of other subjects for Indian school students based on the Mahout machine learning method built on the Hadoop ecosystem. [34] proposed a Hadoop and MapReduce big data collaborative filtering system for resources recommendation such as videos, classes, papers, etc. for the learners based on click streams and web accesses of the users. [23] proposed adaptive course recommendation system based on student profile and teacher’s objective from the class. The system based on genetic algorithm implemented on HDFS and MapReduce on Hadoop big data ecosystem based on analysis of social media data and emotional learner state.

5.1.7. **Instant Feedback:** Big data deployment in e-learning can give the institution and stakeholders instant feedback of which learning style and patterns work and which does not work per each learner [31,35]

5.1.8. **Student Behavior Monitoring:** Provided obtaining the student permission for behavioral data collection (includes data about emails, calls, photos, payments, notes, social media interaction, family, friends, habits, and interests), [36] suggested big data-based system for monitoring the attitudes and behaviors of the student. Behavioral analytics detects the tendency of the student towards violence and terrorism.

6. **Discussion, Challenges and Future Research**

6.1. **Technical Challenges**

Data integration among the data generated by different sources with different structures and formats is a technical challenge for big data analytics in the e-learning platforms. Data collection and handling of data when not all sources are present as missing some data important for processing is another technical challenge that needs to be addressed. The literature in the e-learning big data analytics misses a holistic case study that merges the historical data, behavior data, and other data sources. Yet, acceptance of big data in the e-learning systems is still in a primary phase [26]. The big data analytics in the e-learning is also facing difficulties in providing databases for big data research and analysis in the e-learning field.

6.2. **Big Data Analytics Implementation challenges**

Hadoop has over 190 configurations to affect big data analysis performance. Superdeep and Hadoop configuration tuning need to be investigated to achieve the optimal configuration in the e-learning big data analytics platforms. There is a need for the adoption of big data novel solution by traditional educational institutes. Staff at universities and institutes have to be trained in innovative data management skills. In addition, the cost with the holistic, big data system (storage warehouses, tools, etc.) is still high and to be further optimized and decreased. However, Data standard can be found in
it is not fully applied to the educational data. Data analytics in the education sectors lacks the applicable standard measures and indicators to represent educational variables such as educational progress estimation. Big data analytics researchers should ensure a careful analysis of historical data. Past performance must be used carefully for the only estimation and not for grading. The student may not get under-estimated for previous good or bad performance. Also, similar behavior patterns for several learners may not conclude the similarity between the learning styles and preferences.

6.3. Political Issues
Data privacy and protection, such as the huge information gathered from the learners, can be mistreated and expresses risk for the learners. Data of the learner must have legal protection for its intended use only and must be protected. Also, full consideration of the big learning data and numbers, regardless of the human perspective, represents a challenge. Numbers only as a learning measurement, maybe not sufficient for student assessment.

7. Conclusion
Big data is evolving in the educational sector. Learners and educational organizations are generating large amounts of data to hide important information. The integration of big data technologies in the educational sectors reveals viable information for all involved education process stakeholders. There are several benefits of applying big data in the education field. Educational institutions will benefit from enhanced market analysis, improved adapted learning environment, analyzed career consulting based on student data, minimize students drop-out from classes, support decision making and administration works. Nevertheless, the big data analysis is having several challenges as well, such as data privacy, security constraints, limitation of data integration among different data sources, and the relatively high analysis cost of the holistic, big data solution. Moreover, new future research areas are urging like deployment of other platforms such as Superdoop in the e-learning platforms, standardized data protocols that ensure privacy and security, and holistic case studies that combine historical and different data sources within optimized Hadoop configuration. In summary, e-learning will be boosted with the growing big data penetration in the e-learning.

References
[1] Yanhui W 2018 Language E-learning based on learning analytics in big data era Proceedings of the 2018 International Conference on Big Data and Education pp 106–11
[2] Fee K 2009 Delivering e-learning: a complete strategy for design Appl. Assess.
[3] Kwary D A and Fauzie S 2018 Students’ achievement and opinions on the implementation of e-learning for phonetics and phonology lectures at Airlangga University Educ. e Pesqui. 44
[4] Bhat A Z and Ahmed I 2016 Big data for institutional planning, decision support and academic excellence 2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC) (IEEE) pp 1–5
[5] Santoso L W 2017 Data warehouse with big data technology for higher education Procedia Comput. Sci. 124 93–9
[6] Dougiamas M and Taylor P 2003 Moodle: Using learning communities to create an open source course management system EdMedia+ Innovate Learning (Association for the Advancement of Computing in Education (AACE)) pp 171–8
[7] West D M 2012 Big data for education: Data mining, data analytics, and web dashboards Gov. Stud. Brookings 4 1–10
[8] Alblawi A S and Alhamed A A 2017 Big data and learning analytics in higher education: Demystifying variety, acquisition, storage, NLP and analytics 2017 IEEE Conference on Big Data and Analytics (ICBDA) (IEEE) pp 124–9
[9] Kochetkov O T and Prokhorov I V 2017 The research of approaches of applying the results of big data analysis in higher education AIP Conference Proceedings vol 1797 (AIP Publishing LLC)
[10] Hadoop A 2016 Apache hadoop yarn
[11] White T 2012 Hadoop: The definitive guide (‘‘O’Reilly Media, Inc.’’)
[12] Dean J and Ghemawat S 2004 MapReduce: Simplified data processing on large clusters
[13] Fuad A, Erwin A and Ipung H P 2014 Processing performance on apache pig, apache hive and MySQL cluster Proceedings of International Conference on Information, Communication Technology and System (ICTS) 2014 (IEEE) pp 297–302
[14] Vavilapalli V K, Murthy A C, Douglas C, Agarwal S, Konar M, Evans R, Graves T, Lowe J, Shah H and Seth S 2013 Apache hadoop yarn: Yet another resource negotiator Proceedings of the 4th annual Symposium on Cloud Computing pp 1–16
[15] Ingersoll G 2009 Introducing apache mahout: scalable, commercial-friendly machine learning for building intelligent applications IBM Corp.
[16] Zaharia M, Xin R S, Wendell P, Das T, Armbrust M, Dave A, Meng X, Rosen J, Venkataraman S and Franklin M J 2016 Apache spark: a unified engine for big data processing Commun. ACM 59 56–65
[17] Apache Software Foundation 2018 Apache Spark™ - Unified Analytics Engine for Big Data Apache Spark
[18] Chen J, Tang J, Jiang Q, Wang Y, Tao C, Zhang X and Liao J 2017 Research on architecture of education big data analysis system 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA) (IEEE) pp 601–5
[19] Zhang W and Jiang L 2018 Algorithm Analysis for Big Data in Education Based on Depth Learning Wirel. Pers. Commun. 102 3111–9
[20] Dandan H, Yajuan Z, Junfeng L, Chen L, Mo X and Zhihai S 2017 Research on Centralized Data-Sharing Model Based on Master Data Management MATEC Web of Conferences vol 139 (EDP Sciences) p 195
[21] Liang J, Yang J, Wu Y, Li C and Zheng L 2016 Big data application in education: dropout prediction in edx MOOCs 2016 IEEE Second International Conference on Multimedia Big Data Analysis (BigMM) (IEEE) pp 440–3
[22] Chen G, Wen S, Yang J and Song G 2014 Building an Experimental Platform for Cloud and Big Data Education 2014 International Conference on Education Reform and Modern Management (ERMM-14) (Atlantis Press)
[23] Madani Y, Bengourram J, Erritali M, Hssina B and Birjali M 2017 Adaptive e-learning using genetic algorithm and sentiments analysis in a big data system IJACSA) Int. J. Adv. Comput. Sci. Appl. 8 394–403
[24] Anna N E V and Mannan E F 2020 Big data adoption in academic libraries: a literature review Libr. Hi Tech News
[25] Drigas A S and Leliopoulos P 2014 The use of big data in education Int. J. Comput. Sci. Issues 11 58
[26] Hanapiyah Z M, Hanafi W N W and Daud S 2018 Issues, Challenges and Opportunities of Big Data Management in Higher Education Institutions in Malaysia Indian J. Sci. Technol. 11
[27] Li Y, Li P, Zhu F and Wang R 2017 Design of higher education quality monitoring and evaluation platform based on big data 2017 12th International Conference on Computer Science and Education (ICCSE) (IEEE) pp 337–42
[28] Yu X and Wu S 2015 Typical applications of big data in education 2015 International Conference of Educational Innovation through Technology (EITT) (IEEE) pp 103–6
[29] Denley T 2013 Degree compass: A course recommendation system Educ. Rev. Online 1–5
[30] Zhenyu C 2017 The application of big data in higher vocational education based on Holland vocational interest theory 2017 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII) (IEEE) pp 37–40
[31] Maresca P and Molinari A 2016 Is e-learning ready for big data? And how big data would be useful to e-learning? DMS pp 34–40
[32] AL-Taie R R K and Abu-Alsaad H A 2018 Using Big Data Technology for Prediction of Quiz Difficulty Level in E-learning Systems Iraqi J. Inf. Technol. 8 164–81
[33] Dwivedi S and Roshni V S K 2017 Recommender system for big data in education 2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH) (IEEE) pp 1–4
[34] Shaikh A A Big Data Recommendations Algorithm For E-Learning Using Click Stream Data
[35] Secades V A and Arranz O 2016 Big data & eLearning: A binomial to the future of the knowledge society IJIMAI 3 29–33
[36] Baig A R and Jabeen H 2016 Big data analytics for behavior monitoring of students Procedia Comput. Sci. 82 43–8
[37] Rankin J G 2016 Standards for reporting data to educators: What educational leaders should know and demand (Routledge)