

BIG DATA – OPPORTUNITIES AND CHALLENGES FOR EDUCATION

Valentina Terzieva, Katia Todorova, Petia Kademova-Katzarova

*Institute of Information and Communication Technologies – BAS
valia@isdip.bas.bg, katia@isdip.bas.bg, petia@isdip.bas.bg*

Abstract: *The paper reveals the potential of Big Data applied in education. The specifics of Big Data in educational contexts and different sources for their extraction are described. The power of innovative tools for data collection, management, and analysis by which to identify best practices or problems in the educational process is shown. Considering these findings, a conceptual proposal for an example of educational system using Big Data is presented. The possible effects of the implementation of Big Data in education are discussed.*

Key words: *Big Data, Education, Personalisation, Learning Analytics*

1. Introduction

Recently in today's information society several terms related to new significant technologies have appeared. Along with virtualization and cloud structures, IT professionals and analysts increasingly use the term Big Data. This notion means a huge amount of structured and unstructured information generated at high speed from multiple diverse sources – digital and traditional. It is so large, dynamic and complex that it cannot be handled with conventional applications in real time. The term appears in consequence of the rapid growth of data warehouses, where progress in ICT allows this information to be rationalized and its potential to be exploited in the whole society – economy, health, education, etc.

In education area Big Data are not only "big" in size, but also of huge importance as they allow the analysis and interpretation of a wide range of measurable data, which can be used for better management of an educational process – development of learning and training strategies, assessment of their impact, practical result-based approaches, support of customized teaching, and so on. In our opinion, the use of Big Data in education can lead to a significant improvement of its efficacy and quality. Unfortunately in Bulgaria the implementation is still in the beginning. This article outlines the ways in which Big Data can be used to improve learning and overall educational process and proposes a conceptual frame for educational system using Big Data.

2. Big Educational Data

2.1. The Educational Context of Big Data

Collection, analysis and management of the huge amounts of educational data have to be consistent with their specific characteristics (Figure 1). From an educational viewpoint the term „Big Data” encompasses the following aspects [1]:

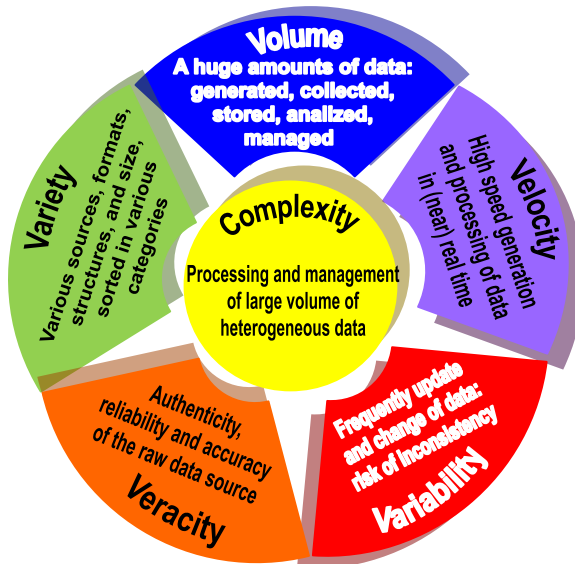


Figure 1. Educational context of Big Data

1. **Volume** – a huge amount of information generated, collected and stored from all kinds of educational sources (schools, universities, organizations, etc.) for all students, courses, learning resources, etc. for specified period of time. It gives a global overview of the whole education sector.

2. **Velocity** – high speed generation and processing of data in (near) real time. All stakeholders require immediate access to the learning data for management of the educational process – learning resources, achievements, assessments, exam results and others. It allows making adequate decisions.

3. **Variety** – a wide range of information generated from various sources in various formats, with various structure and size, sorted in various categories relating to all aspects of the educational process (learning platforms, plans, scenarios, resources, students' learning styles, cognition, portfolios, etc.). It allows preparation of classifications, groupings, correlations.

4. **Variability** – frequently update and change the educational data. Data tracking and validity period of their storage has to be taken into account in order to avoid inconsistency. It requires updating the components of learning process that are used for educational analyzes.

5. **Veracity** – authenticity and reliability of the raw data source, detecting deviations, noise and anomalies. The truthfulness of educational data is ensured.

6. **Complexity** – processing and management of very heterogeneous and large volumes of data from multiple educational sources that are logically combined and analyzed to extract implicitly information. It requires complex and different kind of processing methods.

The upper review of educational aspects of Big Data characteristics gives the reasons for their implementation in educational contexts that is described in details in the next section. On that basis educational strategies can be developed, their impact on all participants in the educational process can be assessed and a personalized approach to learning can be assisted.

2.2. Big Data Sources

The exponentially growing amounts of data in education are generated from both traditional and online sources. Traditional school and academic education produces information mainly concerning personal, demographic and administrative issues. This information presents personal data about a learner (biographical data; training, experience, qualifications and skills; learning style, psychological type; academic achievements, acquired degrees and majors), type of institution, faculty, curriculum plans, study programs and courses, teachers-students ratio, etc. More easily a much larger volume of data is accumulated from online sources during the interaction of students with e-learning platforms. Thus the generated information represents "snapshot" of the participants' behaviour in the learning process. These are current data from the navigation in an interactive learning environment, used learning resources (e.g. text, tests, and exercises), interaction between learners and teachers, cooperation through thematic forums, newsgroups, blogs, vlogs, deposits, chat, social networking, learning games etc. On the other hand, education professionals also generate data – they create, annotate and store learning resources in web-based repositories, build semantic networks and conceptual maps. Furthermore it is also useful to establish and regularly update digital records for all components of the educational process and student profiles. Extraction and processing of such a huge amount of heterogeneous information (unstructured, semi-structured and structured; quantitative, qualitative and textual data) is an extremely complex task that is carried out with mathematical and statistical procedures and methods.

2.3. Educational Data Mining

Initially, data processing occurs as educational task of tracking log files (.log) that contain information about the interaction between users and e-learning platforms [2]. Their analysis provides information, which serves mainly for statistical purposes. Later on the e-learning systems generate increasingly larger volume heterogeneous data of an educational nature which requires the use of other complicated approaches for processing Big Data, based on the well-known data mining methods. For the data retrieval they rely on a variety of tools and techniques for analysis and detection of intrinsic relations, links and patterns in raw data. This helps to reveal existing implicit relationships between the different elements and extraction of essential information necessary for the management and decision making. The methods applied are the same as in statistics, computer science, artificial intelligence, etc. [3]. The term Educational Data Mining (EDM) is used for implementation of techniques for the extraction and processing of a specific type of data that are generated in education field and are crucial to solving educational problems [4, 5, 6]. The most commonly used mathematical methods serve to analyze large sets of data in qualitative and quantitative terms for detection and exploration of functional patterns and trends in the education system processes. Thus answers can be provided to significant questions concerning the characteristics of school environment, learning resources, and programs, how they affect the effectiveness of training, progress and success of learners, etc. Basing on them education professionals and system administrators make informed decisions and build development strategies and predicts for the entire education system.

Usually in research the works methods for processing and management of educational data are classified into two broad categories: 1) statistics and visualization and 2) extraction and processing of information from the local and global networks. The second refers to methods for detecting hidden patterns in large data sets such as clustering and classification, cluster analysis, association rules, neural networks, etc. [7]. The processing of educational data is performed by various mathematical approaches depending on the objectives:

- cluster-, factor-, network- and domain-analysis are used to detect structure and intrinsic characteristics of the data
- correlation-, association rule- and causal data- mining, patterns, templates, help to explore the hidden links and multilateral relationships between the data set elements
- classification, regression, modelling, estimation of density distribution are used for prediction of some unknown aspects of the data item, based on a combination of other known aspects
- visualization of data in an sophisticated readable format allows stakeholders to make well-informed, reasoned judgment.

3. Effects of Big Data Implementation in Education

In the near future it is expected Big Data to help deep understanding of the educational process and related issues. Processing, analysis and visualization of these data are most commonly used for [1, 6, 8, 9, 10]:

Statistics on the students' activities in an educational platform – used learning resources, assessments of completed assignments, tests and exams, interactions with faculty and other students, activity in forums, blogs, vlogs, social networks etc.

Understanding the education process in micro and macro context, in order to identify the most used web resources, their usefulness, distribution of learning activities per time (hours, days, months, etc.), often recommended courses and units, the percentage of the accomplished tasks, the relative difficulty degree of exams, a preferred student's learning style and other available parameters of the educational process that can be researched and analyzed. All these data allow training strategies to be applied successfully.

Tracking the individual learning process of all learners in each e-learning course by analysis of digital traces left behind – gives an idea of the entire learner's educational path, allows the effective determination of individual characteristics, as well as recommendations for future development.

Modelling both the students' profiles (based on initial data and records of learning interactions) and the learning scenarios – they provide information on learning styles of the individual learner and the group as a whole.

Personalization of curriculum courses based on the accumulated data on all aspects of teaching and profiles (portfolios), that serve to meet the individual students' needs and preferences. This increases the effectiveness of the training.

Predicting the academic performance of students based on their personal profiles, digital traces left during interaction with the academic platform and other available data.

Feedback for teachers, authors of learning resources and students. When a resource hinders many learners, the overall perspective on the problem is very useful – to find the exact reason and to get an idea of not only what prevents the knowledge absorption, but also to explore a specific "stumbling block" for the individual learner to correct and/ or to propose alternative methods for overcoming them.

Planning the resources necessary for education in the provided programs and subjects according to the goals, economic strategies, etc., so that to increase the effectiveness of education as a whole.

Collaboration of specialists from various branches in order to support high functionality of the education system promotes cooperation and provides an assessment of multiple aspects.

Motivation for all participants in the educational field to take advantage of the visible results of the appropriate application of Big Data techniques – sooner or later they will find the benefits and convenience using this technology.

The implementation of Big Data in education can radically change it. That is a winning opportunity for both educators and learners. The benefits for teachers are significant and have the potential to affect the future of learning, to make a revolution in the way that it is analyzed and to evaluate its effectiveness. There should be taking into account some risks for not complying with the recommendations of confidentiality, security and reliability.

4. A Concept for Educational System Using Big Data

Worldwide in educational context there are implementations of Big Data mainly in universities and more rarely in schools [9, 10]. Unfortunately, in Bulgaria we have no such practice. Educational data are everywhere today – the problem is that in most cases different detailed key indicators reports are not always readily available, adequate and satisfactory for people who practically need them – school administrators, teachers and other authorities. In our view during the whole learning cycle for each student as early as preschool age data have to be collected in an individual digital portfolio. Those encompass all classroom and extracurricular activities – initial personal information, chronology of training, subjects studied, completed courses, curricula, tests, exams, competitions, achievements, external evaluations, state matriculation examinations, school leaving examinations, obligatory psychological and pedagogical characterization, etc. A substantial part of this information can be extracted from the recently widespread e-learning systems. Already collected, these huge amounts of educational data can be subject to mathematical and statistical analysis to carry out the learning research in micro and macro context.

Data collected at the macro level represent aggregate information about educational institutions, training programs, plans, courses, learning resources, teachers-students ratio, etc. It is used for extracting essential relationships and trends upon which to create or improve educational programs, concepts and policies at class, school, and university level. In particular, it serves for analysis regarding quality and efficiency, strengths and weaknesses, case of problems and finding the best ways to overcome them, support for pedagogical decision making and management.

Particularly important for obtaining a complete picture of the individual learning path is the collected data on micro-level, reflecting the learner's perception style and behaviour in the educational process. They cover his actions in solving problems (an approach, route, speed, quality of decision, trials, mistakes, use of assistance, etc.). Subsequent analysis gives comprehensive information on the personal characteristics of each student, his/ her strengths and problematic sides, which are the basis for the learning resources' adaptation:

- Selection of suitable learning resources – general purpose, copyright and customised according to specific needs, abilities, aptitudes, interests, etc.
- Provision of optional educational and popular science materials, general and specialized web-based libraries, reference books, encyclopaedias, etc.
- Application of individualised teaching and testing methods.
- Preparation of personalized curricula.

Figure 2 shows a conceptual frame for educational system using Big Data: generation, collection and interaction of large amounts of different types of data in the educational process.

All information must be gathered and updated in the *Portfolio* that comprises the *Individual Data* and reflects educational development. They include a priori information, collected personal data and data "snapshots" arising from an on-going problem solving (micro level). The individual data are compared with data of the other students, in order to make summaries and stand out trends, that are the basis for group training according to the interests, abilities, goals, age, etc. – organized courses, clubs, thematic networks, etc.

The data of all students are aggregated in the *Common Data*, where they are classified in categories according to predefined parameters: level of knowledge, skills, motivation, satisfaction and attitudes towards learning process, speed of learning, learning style and preferences, effectiveness of learning environment and resource's usage, demographics, etc.

Big Data use powerful mathematical instruments as *Tools* for different processing methods. The gathered information is processed and analyzed by various mathematical approaches to discover correlations and trends, to make predictions, and to support decision making, which affect the learning process as a whole (macro level). These analyses reveal the development of distinct stages of teaching/ learning processes, which results in a new interpretation of existing information and new knowledge. Particularly in the education can be examined behavioral patterns to identify possible risks and to take appropriate preventive actions. As a result, the teacher can enhance the educational process through appropriate actions.

Most frequently for predicting the students' success the classification and regression methods are used. The goal is to prevent dropping out of university, attend the next academic term and increase the percentage of graduates [9]. Often several methods are implemented simultaneously – quantitative analysis and case analysis, associative rules, clustering, classification, sequence analysis, modelling dependencies, multivariate adaptive regression splines, decision tree, neural networks, etc.

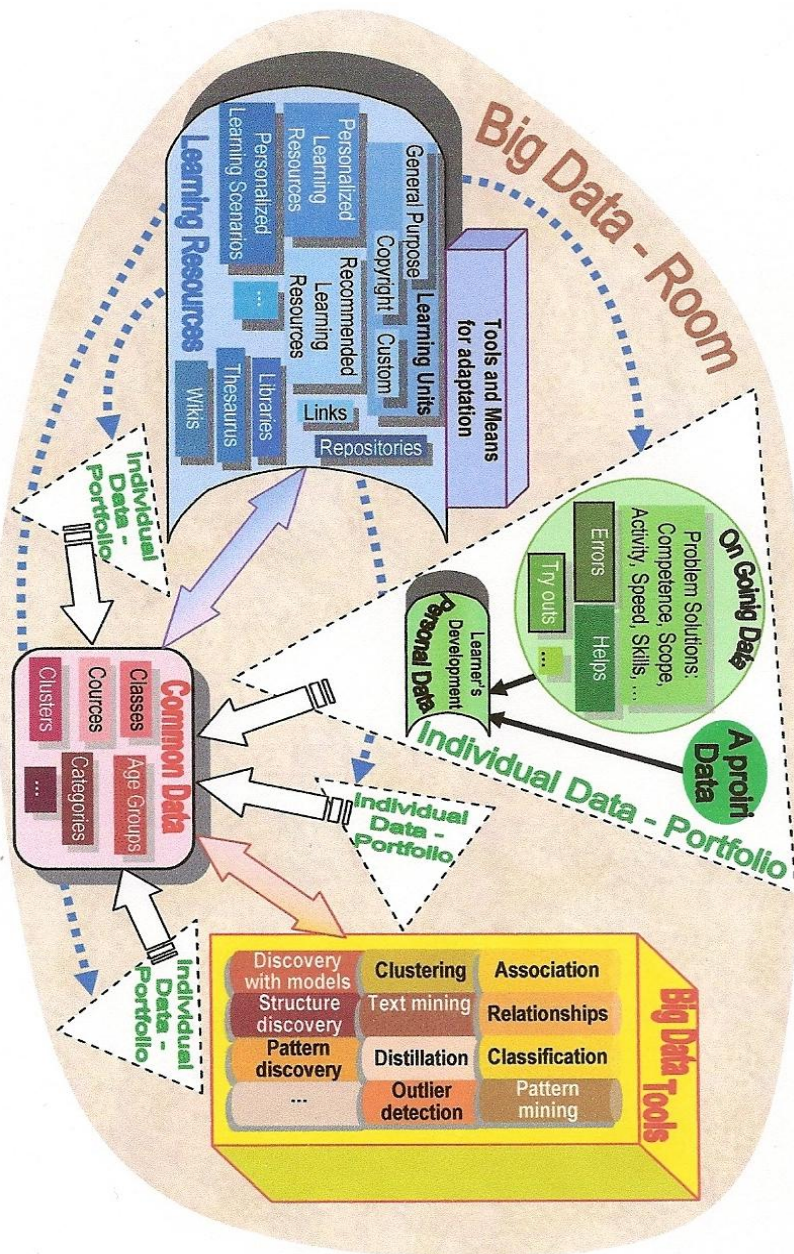


Figure 2. A concept for educational system using Big Data

The resulting summaries and conclusions are used for making policies in education. At the same time personalized *Learning Resources*, plans and scenarios in conformance with specific learner's needs based on the individual data analysis can be offered.

Conclusion

The extraction and accumulation of large amounts of data and their analysis rapidly become ordinary activity in modern society where the efficiency of the education is vital for a successful development. Educational institutions, mainly universities already use tools for processing Big Data in order to improve the services provided and to achieve some predefined values of measurable indicators. At the institutional level such analysis helps detecting areas where an improvement is needed, reporting of results and determination of development policies. The penetration of ICT tools in adaptive learning systems allows exploiting the power of feedback at the individual level. Measurement and visualization of learning activities motivate students to develop skills in monitoring their own learning and to see direct results of their efforts. On the other hand, the teachers receive information about students' achievements in near real time, which helps them to adapt teaching methods or to initiate appropriate interventions in various forms. Big Data tools enable deep understanding of the meaning of collected data and if all the education professionals benefit from them, they can make fundamental changes in the education system.

References

1. Briggs, S., Big Data in Education: Big Potential or big Mistake?, 2014. <http://www.opencolleges.edu.au/informed/features/big-data-big-potential-or-big-mistake/>
2. Bellaachia, A., Vommina, E., MINEL: A Framework for Mining E-Learning Logs, In IASTED International Conference on Web-based Education, Mexico, pp. 259-263, 2006.
3. Jindal, R., Dutta Borah, M., A Survey on Educational Data Mining and Research Trends, International Journal of Database Management Systems, Vol. 5, No. 3, pp. 53-73, 2013.
4. Romero, C., Ventura, S., & García, E., Data Mining in Course Management Systems: Moodle Case Study and Tutorial. Computers & Education, 51(1), pp. 368-384, 2008.
5. Romero, C., Ventura, S., Educational Data Mining: A Review of the State of the Art, IEEE Transactions on Systems, Man, and Cybernetics – Part C, 40(6), pp. 601-618, 2010.
6. Peña-Ayala, Alejandro (Ed.): Educational Data Mining: Applications and Trends, Springer International Publishing, Series Vol. Studies in Computational Intelligence, p. 468, 2014.
7. Baker, R., Yacef, K., The State of Educational Data Mining in 2009: A Review and Future Visions, Journal of Educational Data Mining, 1 (1), pp. 3-16, 2009.
8. Pappas C., Big Data in eLearning: The Future of eLearning Industry, 2014. <http://elearningindustry.com/big-data-in-elearning-future-of-elearning-industry>

9. Luan, J., Data Mining and Knowledge Management in Higher Education - Potential Applications, Annual Forum of the Association for Institutional Research, Toronto, Canada, 2002. <http://eric.ed.gov/ERICWebPortal/detail?accno=ED474143>
10. Arnold, K. E., Pistilli M. D., Course Signals at Purdue: Using Learning Analytics to Increase Student Success, Proceeding LAK '12, pp. 267-270, ACM New York, 2012.

BIG DATA – ВЪЗМОЖНОСТИ И ПРЕДИЗВИКАТЕЛСТВА В ОБЛАСТТА НА ОБРАЗОВАНИЕТО

Валентина Терзиева, Катя Тодорова, Петя Кадемова-Кацарова

*Институт по информационни и комуникационни технологии – БАН
valia@isdip.bas.bg, katia@isdip.bas.bg, petia@isdip.bas.bg*

Резюме: Статията представя потенциала на Big Data, приложени в образованието. Описани са спецификите на Big Data в образователен контекст и различни източници за добиването им. Показана е мощта на иновативните инструменти за събиране, управление и анализ на данни, чрез които да се идентифицират добри практики и проблеми в образователния процес. Като се имат предвид тези констатации, е предложена концепция за примерна учебна система, използваща Big Data. Посочени са възможните ефекти от прилагане на Big Data в образованието.

Ключови думи: Big Data, Образование, Персонализация, Анализ на учебния процес