Proceedings IFTE-2020, 0625-0632
doi:10.3897/ap.2.e0625

VI International Forum on Teacher Education

Identification of Key Factors in the Formation of an Individual Trajectory of Teacher Professional Development in Digital Environment Based on Big Data

Fail M. Gafarov (a), Anis F. Galimyanov (b)*

(a), (b) Kazan Federal University, 420008, Kazan (Russia), 18 Kremlyovskaya street, anis_59@mail.ru

Abstract

The relevance of the study is determined by the need to build new methods for the study of educational activities. Currently, electronic document management has become a daily occurrence and all documents, including diaries, class journals and annual reports, exist in electronic form. These digital data occupy a very large memory and are constantly updated both in volume and in nomenclature. There are special methods for processing unstructured and partially structured data called big data. With proper processing, this data can be used in the further optimal design of the educational process. In this regard, this article is aimed at revealing the features of the use of big data methods in the educational process, as well as the rationale for the use of machine learning methods and neural networks to study the hidden patterns in educational process. The use of big data and machine learning methods will help determine the key factors in the formation of the individual trajectory of the teacher's professional development in the digital environment.

Keywords: big data, educational process, educational environment, neural networks, machine learning.

Published by Kazan federal university and peer-reviewed under responsibility of IFTE-2020 (VI International Forum on Teacher Education)

* Corresponding author. E-mail: anis_59@mail.ru

Introduction

In recent years electronic data storage systems have become commonplace in educational system. All digital everyday data, from the school electronic journal to the electronic reporting materials, is accumulated and subsequently becomes invaluable material for further studies of the effects of various parameters of the educational process on each other. These studies come down to the processing of so-called big data. In essence, these accumulated and constantly updated data are not structured and are not suitable for direct processing. These data must first be structured, stored and processed efficiently. To process a large archives and big data, new technologies are required, which are called Big Data technologies (Nakamoto, 2013; Franks, 2010; Mayer-Shenberger & Kuk'er, 2014). This term refers to a large and complex data sets that can be either structured or unstructured and take up a large amount of memory. Some classification of problems with big data was presented in Chekharin (2016).

Purpose and objectives of the study

The purpose of the study is the disclosure of the features of the use of big data in the educational process, the justification of the need for neural networks to study and apply them to searching hidden patterns of the educational process.

Literature review

Algorithms and programs, created on their basis for process big data, deals with terabytes of data. In practice, these data are heterogeneous, stored in different formats, and as a rule come in at different speeds and from different sources. This data can be processed in a phased analytical pipeline mode, which converts, analyzes and integrates this data. Therefore, the use of artificial neural networks that are capable of processing such data is promising here.

Here it is necessary to especially note the heterogeneity, that is, the multiplicity of formats of incoming and processed data. This factor creates a problem even when there is not too much data. To solve this problem, it is necessary to develop special informational constructions (Tsvetkov, 2014a), to build models of informational interactions (Tsvetkov, 2014b), reflecting the properties of information space Hard Drive Cost As suggested by one website (Hard Drive Cost Per Gigabyte, 2017).

Following Mikhail Leviev (2018), head of AlgoMost, we can distinguish the following main types of data in the field of school education (we have some difference from the traditional approach):

1. Pupil's personal data

2. Data on the effectiveness of training materials

3. Data on teachers, including retraining data.

4. Administrative (system-wide) data

5. Data on the interaction of students with electronic learning systems

6. Predictive data.

Based on the analysis of existing approaches and models, an attempt was made to determine the mega-directions of Big Data (Frumin, 2018):

1) associated with thinking (primarily critical and creative thinking);

2) associated with interaction with others (communication and collaboration);

3) associated with interaction with oneself (self-regulation, reflexivity and self-organization).

But in all existing studies, according to the big data in education, the progress of one of the main participants in the educational process, the teachers, remains obscured. The study of big data in education is valuable because it gives (or should give) material for improvement and regulating the educational process. These measures, as a reaction to negative conditions, are most useful for work on the development of the educational system, including as an integral part of the educational system, for the improvement and professional growth of teachers. The influence of all parameters can only be taken into account by using the methods of neural networks and machine learning. For the effective application of neural networks and deep learning tools to educational data, a significant amount of source data is required for training the model, and in a such situation Big Data methods may be very useful.

Methodology

The expectation of global positive results in educational activities is traditionally assigned and concentrated on the activities of school teachers. This paradigm requires detailed in-depth research and a significant increase in support. The solution of specific pedagogical problems should begin with the formulation and solution of administrative tasks. In this case, informatization could provide effective results if the previous results were adequately taken into account at all levels of the educational process management and optimal

decisions will be made to achieve the ultimate goal. But in practice, decisions are made according to the behavioral principle of “acceptability” by (Herbert, 1965), that is, of all acceptable decisions, not the most effective and optimal for achieving the goal is made, but optimal for the resources used to realize and achieve the goal. This naturally reduces the likelihood of achieving the goal, and possibly distorts it. This principle works in all hierarchical systems, which include the education system.

Currently, information technologies are becoming more and more "friendly" to the user, and there is an erroneous idea that data is easy to obtain. But, in most cases, what a simple user has are not even data, they are only information that needs to be processed on the basis of data states. The ease of obtaining information entails the emergence of redundant information flows, both horizontal and vertical. There are many steps in the hierarchy, and the number of these steps does not decrease, therefore there are also many information flows. This requires special approaches to verify, and to classify, and to aggregate data to extract reliable data from this information, and then to extract knowledge. Traditional methods of periodically extracting and processing only a small part of the results of the educational process, such as the analysis of Unified State Exam (USE) grades, cannot give an objective picture, since the dynamics and trends of the educational process are not taken into account. Therefore, the transition to the management of the educational process based on big data may become an important factor in the development of education. As previously noted, the teacher is a figure on whom the expectation of positive results is based, and therefore the influence of the dynamics of the development of the teacher himself on the dynamics of the educational process as a whole has a great importance. Here it is necessary to exclude the influence of “noise effects” on the main data as much as possible and apply the appropriate mathematical apparatus. This study aims to develop a methodology that allows to identify the most "influential" data and build the appropriate mathematical apparatus.

Research Methods and Techniques

There are several factors contributing to the intensive introduction of the big data methods in the field of education. Significant reduction in the cost of storage of a unit of information plays here an important role. The level of development of information storage technologies leads to a decrease in the cost of storing one gigabyte of information by 46% over the past 9 years, and over the past two decades - by more than 250% as suggested by one website (Hard Drive Cost Per Gigabyte, 2017). The significance of these changes leads to the fact that the majority of information is stored only in digital form. Educational organizations have almost intensively use electronic document management systems, as well as to saving the maximum possible amount of information about interaction with their partners and students. New software and hardware, which are emerging and being implemented makes possible to obtain new knowledge from

arrays of information that, more recently, would be considered ballast and destroyed. Today, analyzing various data about students, classifying them and processing them accordingly it is possible to build more productive educational relationships and optimize the educational process. A number of scientists notes, that databases are no longer just a way to store data. They turn into a modern tool for acquiring new knowledge, thanks to which the processes of adaptation and collaboration become more efficient, and decisions made in real time become more personalized, taking into account the specifics of specific participants in the relationship (Bollier & Firestone, 2010). Therefore, the huge amounts of data accumulated earlier, when used appropriately, can become a source of information that will be the key to high educational standards for all students, regardless of personality characteristics and needs.

Based on the foregoing, the methodological basis of our study is the formalization of big data, aimed at the development of the educational system through the identification of patterns in the education system.

There are distinguishing features of Big Data in education from other samples. They can be described by the 5V rule:

1V (volume): the amount of physical data is significant. For example, all exam grades located in a single database.

2V (velocity): the speed of data collection and processing speed of the results is relatively high. For example, grades are saved to database according to a schedule.

3V (variety): variability of processing algorithms for various types of collected results. For example, grades for students in a school can be presented in different sections (gender, age, teacher's person, etc.)

4V (veracity): high reliability of the data collected, allowing to formulate representative results. For example, assessments of the exam, the results of the republican tests.

5V (value): the value of the accumulated data should be concluded in the possibility on the basis of them to formulate useful diverse aspects of the education system. For example, a high level of correlation by class.

To structure the Big Data management processes in education, five interrelated groups of processes can be distinguished (Utomov & Gorev, 2018):

1) Goal-setting - determining the purpose and objectives of the study;

2) Planning - selection of information sources, data acquisition procedures, information processing algorithms;

3) Data collection- organization of data collection in a single database;

4) Analysis of indicators- analysis of the data obtained, determining ways of presenting the results;

5) Adjustment- development of practical regulatory measures;

6) Completion- regularities detection.

Results

As noted earlier, the methodology of our study is to formalize the handling of big data, obtained as a result of the educational process. A large amount of data allows us to find more accurately various relationships for further presentation of analytics in an aggregated, readable form. The variability of the data allows to identify hidden dependencies where at first glance they are not worth looking for. For example, the dependence of the grades of the test on the day of the week.

The processing speed of Big Data information is close to real time. First of all, data is collected on storage facilities (local or network), which can be either disparate or combined into a single system. Information is necessarily duplicated to exclude possible losses and it is characterized by a lack of structure, i.e. it can be text, images in various formats, voice, music, etc. Subsequently, the data is processed by an algorithm written by programmers to obtain information in a form convenient for users.

After bringing the data to a formal form, the patterns in the data are searched through deep learning of the neural network. The result of the work is a trained neural network that allows you to build a forecast based on the declared input data. For data smoothing, the Lagrange interpolation polynomial is used (Hussien, 2011).

Discussions

The application of the big data methodology in education for a wide range of tasks essentially remains a problem that has not been fully explored, and probably cannot be fully explored. Because each subject of the educational process is unique and it is hardly possible for everyone to apply common approaches without exception. Therefore, most likely, it is necessary to carry out a deeper classification of data types

and component-wise decomposition of students' achievements. Of course, there are some general trends, and we are at the very beginning of the path, trying to identify them.

Conclusion

The article discusses the methods of processing big data in educational systems. With proper processing, this data can be useful in the further optimal design of the educational process. In this regard, this article reveals the features of the use of big data in the educational process, substantiates the need for the use of neural networks to study and apply the hidden patterns of the educational process. It is proposed to use an interpolation polynomial to smooth the initial data.

Acknowledgements

The reported study was funded by RFBR, project number 19-29-14082.

References

Bollier, D., & Firestone, C. M. (2010). *The promise and peril of big data. Communications and Society Program*. Washington: Publisher.

Chekharin, E. E. (2016). Big data: big problems. *Prospects for Science and Education*, *3*(21), 7-11.

Franks, B. (2010). *Taming the big data tidal wave*. City: John Wiley & Sons, Inc.

Frumin, I. D. (2018). Trends in the development of the content of education: key competencies and new literacy. *Materials of the IV International Forum on Pedagogical Education.* http://ifte.kpfu.ru/ru/lectures/trendy-v-razvitii-sod.

Hard Drive Cost Per Gigabyte (2017). BackBlaze. https://www.backblaze.com/blog/hard-drive-cost-per-gigabyte.

Hussien, K. A. (2011). The Lagrange Interpolation Polynomial for Neural Network Learning. *International Journal of Computer Science and Network Security*, *3*(11), 255-261.

Mayer-Shenberger, V., & Kuk'er, K. (2014). *Big data. A revolution that will change the way we survive, work and think.* Moscow: Mann, Ivanov i Ferber.

Nakamoto, P. (2013). BIG DATA: The revolution that is transforming our work, market and world. Data Analysis / Kindle Edition.

Simon Herbert, A. (1965). *Administrative Behavior*. New York: Free Press. http://www.google.ru/books?id=jmzWLn8pBKUC

Tsvetkov, V. Ya. (2014 a). Information Constructions. *European Journal of Technology and Design*, *3*(5), 147-152.

Tsvetkov, V. Ya. (2014 b). Natural and artificial information field. *International Journal of Applied and Fundamental Research*, *5*(2), 178-180.

Utomov, V. V., & Gorev, P. M. (2018). Development of educational systems based on Big Data technology. *Scientific and methodical electronic journal "Concept"*, 6, 449–461. http://e-koncept.ru/2018/181039.htm.