ATLANTIS PRESS

Advances in Social Science, Education and Humanities Research, volume 437

International Scientific Conference "Digitalization of Education: History, Trends and Prospects" (DETP 2020)

Application of Big Data in Determining and Regulating Trends in Education

Galimyanov A.F.* Gafarov F.M. Muzafarova A.I.

Kazan (Volga region) Federal University
**Corresponding author. Email: anis_59@mail.ru*

ABSTRACT
The relevance of the study is determined by the need to build new methods for the study of educational activity, which is aimed at developing competencies. Competencies are considered as functions of ABC-abilities. As an example, the expression of mathematical culture is given as a vector function of ABC-abilities. It is necessary to process a large amount of data obtained as a result of such formalization to determine and regulate the trends of the educational activity. These data occupy a very large memory amount, are constantly updated both in volume and in nomenclature. These possibly unstructured data are called big data and special methods for working with them are created. This data can be useful in the further optimal design of the educational activity with proper processing. In this regard, this article is aimed at revealing the features of big data application in the educational process, substantiating 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.
Keywords: *ABC-abilities, competencies, big data, neural networks*

1. INTRODUCTION

1.1 Relevance of the Problem

The educational situation that has developed in recent years in Russia and around the world determines the need to rethink the main methodological approaches to the practice of making and implementing those decisions that are related to education and training in schools and further education in accordance with the requirement of time. In connection with the transition to a competency-based approach, it is necessary to identify those key competencies, on the basis of which it is possible to determine and develop competencies for certain requirements. Key competencies for lifelong learning – The European Frameworks [1] defines competencies as a combination of knowledge, skills, and relationships in an appropriate context. Frameworks have established eight key competencies. "Mathematical Literacy and Basic Competencies in Science and Technology" are in the third line of this list. In our opinion, this shows the importance and necessity of mathematical education both in school and in further education. In this regard, it is important to determine the mathematical culture of the individual and the components of mathematical culture. The main trends in education and the ways of its modification can be determined on their basis. To process such data, the use of the big data mechanism is proposed. This is due to the fact that almost all everyday data, starting from the school electronic journal to the reporting materials, are accumulated and subsequently become invaluable for further studies of the effects of various parameters of the educational process on each other. So far, these data are constantly accumulating, but they are not structured and cannot practically be processed mathematically. First data should be structured it in one way or another to process such data. Since there are so many of them, archives are large and constantly updated, it is necessary to apply new technologies, which are called Big Data technologies [2,3,4]. This term refers to large and complex data sets, which can be either structured at this moment or unstructured and take up a large amount of memory. Some classification of problems with big data was presented in [5]

1.2 Analysis of Existing Results

To determine the trends in education, the most important part of which is mathematical education, it is necessary to introduce one or another metric based on which quality monitoring can be carried out. We rely on the definition of quality monitoring according to the methodology of Professor Nuriev N.K., which is described in [6]. According to them, competence (as the ability to solve any problems) in any field is invariantly supported by a triad of abilities <A, B, C> of a certain level of development, i.e. ABC-abilities and interiorized knowledge, as auxiliary means. Here A is formalization ability, B is constructive ability, C is performing ability. The vector, algorithms, and programs created on their basis that should process (F1 (A), F2 (B), F3 (C), F4 (A, B) can be determined based on them, following [6]. F5 (A, C), F6 (B, C), F7 (A, B, C)), which will reflect the student's mathematical culture to one degree or another. The construction of the components of this vector can be done in different ways, based on the desired goal and on the basis of the specific formed competence. In

turn, basic abilities can also be measured in various ways. Then a large amount of data is obtained. To identify trends, it is necessary, for example, to cover data for a whole region. Then we get big data that deals with data calculated in terabytes. We receive the data, most likely heterogeneous, and they come in different formats, and as a rule, at different speeds, and from different sources. You can process this data in a mode of phased analytical conveyor, which converts, analyzes, and integrates this data. Therefore, the use of artificial neural networks that are capable of processing such data has potential here.

Here it is necessary to especially note the heterogeneity, that is, the multiplicity of formats of incoming and processed data. This in itself creates a problem even when there is not too much data. To solve this problem, it is necessary to develop special informational constructions [7], build models of informational interactions [8], reflecting the properties of information space [9].

Following Mikhail Leviev, head of AlgoMost (http://www.edutainme.ru/post/big-data-edu/), we can distinguish the following main data types in the field of school education (they differ from the traditional approach):

1. Student's personal data
2. Learning performance data
3. Teacher's data, including data on retraining.
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 [10]:

1. associated with thinking (primarily with 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 the progress of one of the main participants in the educational process – teacher – remains obscured in all existing studies on the big data in education. The study of big data in education is valuable because they give (or should give) material for improvement, for regulating the educational process. Measures aimed at improving the qualifications and retraining of teachers, as a result of which teachers will be able to form the necessary competencies for their students, are very important in terms of the formation of competencies. 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. This method was applied since the influence of all parameters can only be taken into account with the use of the apparatus of neural networks. The fact is that for the effective operation of neural networks and deep learning tools, a significant amount of source data for teaching the model is required. In such a situation, you can not do without Big Data tools.

2. METHODS/METHODOLOGICAL GROUNDS

2.1 Study Purpose

State standards aim at the formation of well-defined competencies among students. In turn, these competencies are the function of a very large number of arguments, the influence of which on the final result must be determined. The work considers the formation of mathematical culture, as the most important of the components of the learning results, in turn, it consists of smaller components. These components, in turn, affect the formation of well-defined competencies. In our opinion, this approach can be distributed to almost all subjects. But teaching staff is the main actors. The expectation of global positive results in educational activities is traditionally assigned and concentrated on the activities of teaching staff, that is, school teachers. This paradigm requires detailed in-depth research, and if necessary, a significant strengthening of the support. That is, the solution of specific pedagogical tasks should begin with the formulation and solution of managerial tasks. Informatization here could provide effective results if the previous results were adequately taken into account at all levels of the educational process management and optimal decisions were made to achieve the ultimate goal. But decisions are made in practice according to the behavioral principle of "acceptability" by G. Simon [11], 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 probability of achieving the goal, and possibly distorts it. This principle works in all hierarchical systems, which include the education system.

Currently, information technology is becoming more and more "friendly" to the user, an erroneous idea appears that data is easy to receive. But, in most cases, what a simple user has is not even a data, it is only information that needs to be processed to the state of the data. 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, that is why there are also many information flows. And this requires special approaches to verify, and to classify, and to aggregate data to extract from this information reliable data, and then knowledge. Traditional ways of periodically extracting and processing only a small part of the results of the educational process, such as the analysis of Unified State Examination 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 here an important factor in education development. As it was previously noted, the teacher is a figure on whom the expectation of positive results rests, and therefore the influence of the dynamics of the development of the teacher himself on the dynamics of the educational process as a whole is of great importance. Here it is necessary to exclude to the maximum the influence of "noise effects" on the main data and apply an appropriate mathematical apparatus. The study aims at developing a methodology that allows to identify the most "influential" data and building the appropriate mathematical apparatus.

2.2 Methods and Methodologies of the Study

To monitor trends in the educational process, it is necessary to distinguish composite (ABC)-abilities in all subjects, as it was done in this work for mathematical culture, then on their basis to define as functions, competencies. Then we get several data groups. It is necessary to apply the big data mechanism for their analysis. There are several factors contributing to the massive introduction of big data analysis in the field of education. Significant reduction in the cost of storage of a unit of information is essential. The level of development of information storage technologies has led 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% [9]. The significance of these changes has led to the fact that today the vast majority of information is stored only in digital form. Educational organizations have almost switched to electronic document management, as well as to preserving the maximum possible amount of information about interaction with their partners and students. New software and hardware, which are emerging and being implemented at a very fast pace, make it possible to obtain new knowledge from arrays of information that, more recently, would be considered as ballast and destroyed. Today it is possible to build more productive educational relationships and optimize the educational process by analyzing various data about students, classifying them, and processing them accordingly. A number of scientists note 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 [1]. Therefore, the colossal 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 above, the methodological basis of our study is the formalization of big data operating, aimed at developing competencies based on the analysis of ABC-abilities through identifying 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 volume of physical data is significant; for example, all Unified State Examination grades are in the same database;

2V (velocity): the velocity of data collection and the velocity of processing of the results is relatively high; for example, grades are made according to the schedule;

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

4V (veracity): high veracity of the data collected, allowing to formulate representative results; (for example, Unified State Examination grades, Republican Examination Results)

5V (value): 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.

Five interrelated groups of processes can be distinguished for structuring the Big Data management processes in education [12]:

1. goal-setting: determining the purposes and objectives of the study;
2. planning: the selection of information sources, data acquisition procedures, information processing algorithms;
3. data collection: the organization of data collection in a single database;
4. analysis of indicators: the analysis of the data obtained, determination of methods for results presenting;
5. adjustment: the development of practical regulatory measures;
6. conclusion: fixing patterns

3. RESULTS

As it was noted earlier, the methodology of our study is the formalization of the formed competencies, which we showed on the example of mathematical culture, as well as the formalization of big data operating obtained as a result of the educational process. A large amount of data allows to more accurately find various connections for further presentation of analytics in an aggregated, readable form.

The variability of the data allows to identify dependencies where at first glance they should not be. For example, the dependence of the test grades on the day of the week or the dependence of the ability to abstract mind from the living place.

The processing speed of information in big data is close to real time.
First of all, the data is collected on datastores (local or network), which can be either segmental 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 humans.
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 to make a prediction based on the declared input data. The Lagrange interpolation polynomial is used for data smoothing [13]

4. CONCLUSION

The article discusses the methods of processing big data that are obtained in the process of formalizing competencies by ABC abilities. With proper processing, these data can be useful in the further optimal design of the educational process for determining the further development trend in one direction or another. In this regard, this article reveals the features of big data application in the educational process, substantiating the need for the use of neural networks to study and apply the hidden patterns of the educational process. It was proposed to use an interpolation polynomial to smooth the initial data.

ACKNOWLEDGMENT

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

REFERENCES

[1] Bollier D., Firestone C. M. The promise and peril of big data // Communications and Society Program. Washington, 2010. 1

[2] Pat Nakamoto. BIG DATA: The revolution that is transforming our work, market and world. Data Analysis / Kindle Edition, 2013. P. 212. 7

[3] Franks B., Taming the big data tidal wave. John Wiley & Sons, Inc. 2010. P. 341. 3

[4] Mayer-Shenberger V., Kuk’er K. Bol’shie dannye. Revolyutsiya, kotoraya izmenit to, kak myzhivem, rabotaem i myslim. Moscow: Izdatel’stvo«Mann, Ivanov i Ferber», 2014. P. 208. 6

[5] E.E. Chekharin. Big data: big problems. Perspektivy Nauki i Obrazovaniya. pnojournal.wordpress.com/archive16/16-03 Data publikatsii: 1.07.2016 No. 3 (21). S. 7-11. 2

[6] Galimyanov A. F., Talal N. M. A web application for adaptive learning on the basis of ABC-abilities //Journal of Fundamental and Applied Sciences. – 2017. – T. 9. – No.. 1S. – C. 1859-1870. 14

[7] Tsvetkov V. Ya. Information Constructions // European Journal of Technology and Design, 2014, Vol (5), No. 3. p.147-152. 9

[8] Tsvetkov V. YA. Yestestvennoye i iskusstvennoye informatsionnoye pole// Mezhdunarodnyy zhurnal prikladnykh i fundamental'nykh issledovaniy. – 2014. – No.5, ch.2. – S. 178–180. 13

[9] Hard Drive Cost Per Gigabyte // BackBlaze. URL: https://www.backblaze. com/blog/hard-drive-cost-per-gigabyte (data obrashcheniya: 07.11.2017). 4

[10] Frumin I. D. Trendy v razvitii soderzhaniya obrazovaniya: klyuchevyye kompetentsii i novaya gramotnost' // Materialy IV Mezhdunarodnogo foruma po pedagogicheskomu obrazovaniyu. – URL: http://ifte.kpfu.ru/ru/lectures/trendy-v-razvitii-sod. 12

[11] Simon Herbert A. Administrative Behavior. New York: Free Press, 1965. 368 p. URL: http://www.google.ru/books?id=jmzWLn8pBKUC 8

[12] Utomov V. V., Gorev P. M. Razvitiye obrazovatel'nykh sistem na osnove tekhnologii Big Data // Nauchno-metodicheskiy elektronnyy zhurnal «Kontsept». – 2018. – No. 6 (iyun'). – S. 449–461. – URL: http://e-koncept.ru/2018/181039.htm. 11

[13] Khalid Ali Hussien. The Lagrange Interpolation Polynomial for Neural Network Learning. IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.3, March 2011. 5