Analyse the Influence of Big Data on Students’ Learning Behavior

Jumei Ai ¹, Jinlong Gao ², and Ping Du ¹

¹Jiangxi Engineering Laboratory on Radioactive Geoscience and Big Data Technology, East China University of Technology, Nanchang, Jiangxi, China
²Fuzhou Teachers’ College, East China University of Technology, Fuzhou, Jiangxi, China

jmai@ecit.cn, lgao@ecit.cn, pdu@ecit.cn

Abstract. The application of big data has penetrated into many fields of society, and the education industry is no exception. In this paper, we try to collect and analyze the data related to teaching activities, use data mining technology to construct a mathematical model of the corresponding large teaching data, explore the relationship between representation and substance, provide useful information support for teaching activities, promote and help improve the quality of teaching, and verify the proposed mathematical model through experimental analysis. Better analysis results.

1. Introduction

Big data is a high-tech development, developed technology, and a high-tech era of information circulation. It refers to a collection of data that cannot be captured, managed, and processed by conventional software tools within an affordable time frame. The big data is characterized by a large amount of data and a large amount of information, and its value is also large. Effectively collecting and analyzing big data resources of all sizes and using more effective ways to tap the huge value hidden by big data has become a hot spot for many researchers[1].

Big data has spawned many changes in many areas of society. Similarly, in the field of education, the application of big data is very frequent. Domestic scholars have many research results in big data on teaching state data and teaching quality monitoring, innovation and intelligent change of smart higher education teaching content based on big data, and how data processing and services are combined with big data and cloud computing[2]. World-renowned universities such as Yale University, Harvard University, and Stanford University have also launched educational big data-related research programs; American Association of School Administrators (AASA) and School Network Association (COSN), and global information technology research and consulting firm Gartner A project called “Closing the Gap: Turning Data into Action” was implemented to promote the use of big data in student information systems and learning management systems[3].

The teaching process will be influenced by many factors of teaching and learning. Try to collect and analyze the data related to teaching activities, use data mining technology, construct corresponding mathematical models, explore the relationship between appearance and essence, and provide useful information support for teaching activities. The development of supplementary teaching, promoting and helping to improve the quality of teaching, this is the problem we want to explore.

Although the application of big data has penetrated into higher education, there is not much research on the systematic analysis of the impact of big data on the process of teaching activities in colleges and universities. Firstly, based on literature and theory, this paper explores the dynamic relationship and specific action points of big data and teaching activities, and constructs the logic of
the influence of big data on the teaching process. Secondly, it studies the process of big data directly acting on teaching activities, trying to use the mechanism and the two major modules of influence content to summarize the theory of the impact of big data on the teaching process.

This paper analyzes the impact of big data on college teaching process, follows the principle of higher education to cultivate talents as the core, and focuses on the application of big data technology in teaching activities, including big data on teaching design, teaching activity process and teaching management. The impact of the proposed new trend of big data technology in higher education teaching activities, to provide new technology application ideas for improving teaching quality.

2. Collection of College Students’ Learning Behavior Data

Through the school educational management system, examination system, supervision feedback, teachers’ usual attendance records and other channels to obtain data related to teaching activities as a resource [4]. For example, taking the college students’ learning attendance record as their usual performance, the research on the relationship between learning attitude and academic achievement is studied and discussed.

Starting from the students’ usual performance, through the large number of student registrations recorded by the school, one of the students’ scores is analyzed to see if there is a certain correlation. The attendance data recorded by the school is extracted, and the corresponding final grade information is also extracted. Assume that the extracted information has two kinds of information: attendance data and final grade. Among them, the attendance data is processed and the attendance rate is generated. Each student will have a number of courses, each of which has a final grade for the course and the attendance rate of the student’s usual participation in the course. According to the data of each part, the corresponding relationship is analyzed as shown in Figure 1.

![Figure 1. Student Course E-R](Image)

After the attendance information is extracted from the attendance database, the following attendance table is obtained:

| SNumber   | Cnumber   | Mark | Attendance |
|-----------|-----------|------|------------|
| 162050401 | 2204011201| 88   | 0.97       |
| 162050401 | 2204011202| 80   | 0.95       |
| 162050401 | 2204011203| 94   | 1          |
| 162050401 | 2204011203| 89   | 0.98       |
| 162050402 | 2204011201| 69   | 0.9        |
| 162050402 | 2204011202| 67   | 0.9        |
| 162050402 | 2204011203| 77   | 1          |

……
3. Construct Mathematical Models Related to Learning Data and Learning Behavior

Through the acquisition, management and analysis of educational big data, construct relevant mathematical models to analyze students’ learning behaviors, and predict students’ future learning situation. They can also guide students’ academic planning, life goals, and assist teachers in adjusting[5]. Teaching methods, teaching progress, etc.

Big data mining methods include classification, regression analysis, clustering, association rules, neural network methods, and Web data mining. These methods mine data from different perspectives. This project intends to use regression analysis to conduct research.

3.1. One-dimensional Linear Regression Model

In the process of scientific research, we notice the influence of various factors on a certain index, and hope to select some factors, and the effect of the index is ideal. For example, in the treatment of Chinese medicine, the composition of the drug has a greater impact on the condition, and which is the most favorable for the improvement of the condition. It is necessary to study a better formula through a large number of "experiments". There is a dependent variable and a series of independent variables that affect this dependent variable. This requires the application of regression analysis to deal with this series of data problems. The method of regression analysis was first proposed by the 19th century British scientist and explorer Francis Galton. The regression analysis includes unary linear regression, multiple linear regression, stepwise regression, successive regression, and probability regression.

\[
\begin{align*}
  y &= a + bx + \varepsilon \\
  \varepsilon &\sim N(0, \sigma^2)
\end{align*}
\]

Equation 1 is a linear regression model, where \(a\) is a constant term coefficient, \(b\) is a regression coefficient, and a straight line \(y=a+bx\) is a regression line equation. \(\varepsilon\) is a random error, and it is generally assumed that it follows a normal distribution, that is, \(x\) can be a non-random variable, called an independent variable, and \(y\) is a random variable called a dependent variable. Here we need to find the values of \(a\) and \(b\) by the least squares method to ensure that the function of \(y=bx+a\) best matches the \((x_1, y_1), (x_2, y_2), (x_3, y_3), \ldots, (x_n, y_n)\) these data. The least squares method (also known as the least squares method) is a mathematical optimization technique that finds the best function match of the data by minimizing the sum of the squares of the errors. After a series of calculations, an equation can be obtained by the least squares method:

\[
\begin{align*}
  \sum_{i=1}^{n} x_i^2 &\quad b = \frac{\sum_{i=1}^{n} x_i y_i - n \bar{y}}{\sum_{i=1}^{n} x_i^2} \\
  \sum_{i=1}^{n} x_i &\quad a = \frac{\sum_{i=1}^{n} y_i}{n}
\end{align*}
\]

With Equation 2, you can calculate the values of \(a\) and \(b\) to determine the empirical formula \(y=bx+a\).

3.2. Establish Data Association and Analysis

By experimenting with multiple sets of teaching-related data into mathematical models, test and explore the correlation between data. For example, students’ grades in school are good or bad. What causes the students’ grades directly or indirectly, the students’ learning attitude can be reflected indirectly from their study attendance, and the usual attendance data can be obtained. Reflecting students’ learning attitudes also indicates that there is a correlation between learning attitudes and academic performance. Therefore, it is of practical significance to explore the correlation between daily attendance and student achievement.

According to the above attendance rate table 1, the data is plotted as a scatter plot as shown in Figure 2:
Randomly take a set of x, y data from Figure 2 and substituting into Equation 2 to find a and b. Assumption:
\[ X_i = (10, 20, 30, 40, 50, 60, 70, 80, 90, 10) \]
\[ Y_i = (0.56, 0.6, 0.67, 0.75, 0.81, 0.85, 0.90, 0.93, 0.96, 1) \]
Then there are:
\[
\begin{align*}
38500b + 550a &= 483 \\
550b + 10a &= 8.03
\end{align*}
\]
Solved: \( b = 0.005 \), \( a = 0.528 \)
Therefore, the formula \( y = 0.528 + 0.005x \) is obtained, that is, the score x and the attendance rate y roughly satisfy the linear relationship. Of course, the larger the test data amount, the higher the accuracy rate.

4. Conclusion and Thinkings
This result shows that there is a linear relationship between student attendance and academic achievement, with the exception of individual exceptions. If the selected big data angles are different, the impact on teaching activities is different. Subsequent research can continue to select the relevant data of the teacher lectures to refine and process, and find out the connection between them and the teaching effects.

The field of education big data is in the ascendant. It has enormous potential for development and faces many challenges. These challenges include data standards, data collection, model building, data services, open sharing, and privacy protection. The United States attaches great importance to the establishment of educational data standards to regulate the expression and processing of data. The United States Department of Education has organized relevant agencies to develop the Universal Standard for Educational Data, and it has been continuously adjusted and optimized in the process of application. It has developed to the fifth edition in 2015. The data model includes 10 domains: preschool education, primary and secondary education, vocational education, adult education, career and technical education, on-the-job promotion, learning standards, learning resources, evaluation, identification and empowerment\(^6\). China’s education big data field is in its infancy. Its development needs the deep integration of big data technology and education field. Education authorities need to promote the formulation of relevant laws and regulations on education big data, delimit boundaries, clarify responsibilities and powers, establish more comprehensive standards for education data for the national level. Lay the foundation for large scale data sharing and analysis. The educational institutions represented by schools need to enhance the awareness of data-driven teaching and management, build a comprehensive data acquisition environment, and establish a data management
and application mechanism[7]. Research institutions need to integrate educational science with data science closely, carry out multi-disciplinary collaborative research, and focus on the transformation of research results. To jointly promote the healthy development of educational big data and better serve the educational cause.

5. Acknowledgements
This work was financially supported by Jiangxi Engineering Laboratory on Radioactive Geoscience and Big Data Technology(JELRGBDT201805),and University education reform research project of Jiangxi (JXJG-17-6-2 - Analysis of the influence of big data on teaching), and Natural Science Foundation Project of Jiangxi (2015ZBAB201001-Research on private information protection strategies and algorithms in social networks ).

6. References
[1] Zhang Yingchao. Study on Big Data to Influence the Development of Higher Education [D]. Chongqing University. 2016:2-10.
[2] Qian Fenghuan. The Collection of Basic State Data of University Teaching Under the Background of “Big Data” [J].Journal of Wenshan University, 2015 (04): 81-83.
[3] Zhang Yu, LI Yue. Learning Analytics and Education Assessment Based on Big Data Generated by MOOCs [J] Tsinghua Journal of Education, 2013(04):22-26.
[4] The Four V’s of Big Date[EB/OL]. http://www-01.ibm.com/software/data/bigdata/ what-is-big-data.html. [2014-04-19].
[5] Benjamin S. Shifting from Data TO Evidence For Decision Making[J]. Phi Delta Kappan, 2014, 95(7):45-49.
[6] Zheng Yanlin, Liu Haimin. Path Analysis of the Application of Big Data for Education Evaluation in the United States. Theory of Educational Technology [J].2015.07, 342rd.28-29.
[7] Sun Hongtao, Zheng Qinhua. The Technology, Application And Trends of Educational Big Data. Journal of distance education[J]. 2016.05.04, 47-48.