

# Analysis and Decision-making Discussion on the Impact of Educational Big Data in Local Colleges and Universities

Jing Dong<sup>1,a</sup>, Bo Cheng<sup>2,b</sup>

<sup>1</sup> School of Information Engineering Qujing Normal University Qujing, China

<sup>2</sup> School of Marxism Studies Qujing Normal University Qujing, China

<sup>a</sup> Dongjing@mails.qjnu.edu.cn, <sup>b</sup> 916703339@qq.com

**Keywords:** Educational big data; Local colleges and universities; Data mining; Intelligent learning

**Abstract:** In order to achieve intelligent learning, intelligent evaluation and intelligent services, it's important to utilize machine learning, mathematical statistics and neural network data mining technology for dealing with cross-border integration of university education status data, educational management departments and third-party related data. Firstly, this paper analyses the impact of educational big data on the teaching management of higher education as well as the impact of educational big data on the development of higher education; secondly, it introduces the data mining technology which is widely used at present; finally, it illustrates the value and broad application prospects of educational big data by mining the specific case of big data about students' employment.

## 1. Introduction

With the popularization of information and communication technology and the continuous maturity of data mining technology, the era of big data has come (1). The application of big data in the field of education has also begun to attract the attention of many organizations and professionals. Especially in Colleges and universities, students' learning styles have become more diversified. If they follow the traditional standardized subject settings, teaching content, teaching methods, and examination system and personnel training mode, it is easy to ignore students' personality characteristics and cognitive development (2,3). Big data of education record and quantify students' learning situation. If huge data of user's learning behaviors are analyzed from huge educational data, such as the establishment of self-adaptive learning system to evaluate and explain the important factors affecting learning effect, teachers can effectively integrate scattered teaching resources, truly understand each learner (4), and according to their personal characteristics, such as learning habits and learning. Learning interest and learning preference provide accurate and targeted teaching; at the same time, provide learners with personalized adaptive feedback, so that each learner can tap out their own learning preferences, recommend courses and textbooks that best match their learning plan for learners, achieve the best learning effect, and achieve "wisdom" learning (5,6). Through cross-border data integration and cross-border mining (7), it can improve the backward traditional teaching relationship and achieve more personalized interaction.

Employment is top priority in the work of colleges and universities. Mining educational big data can realize the intelligent management of educational equipment and assets, the improvement of students' employment and assistance system, the precise analysis and recommendation according to students' employment intention; big data can also realize intelligent scientific research, through the analysis of big data of scientific research (8,9), the innovation of scientific research procurement mode can be promoted, and the quality of literature service can be improved. The inappropriate scientific research management mode should be promptly warned and transformed.

## 2. The Impact of Educational Big Data on the Development of Higher Education

The impact of big data on the development of higher education is manifested in the following

three aspects: educational big data brings about the openness and diversification for higher education; data mining and analysis technology brings about the individualization for higher education; big data promotes the popularization of higher education.

### **2.1 Educational Big Data Brings about the Openness and Diversification for Higher Education.**

It embodies the openness of teaching resources and learning environment in higher education (10). Colleges and universities not only have abundant teaching resources, but also can provide resources to network operators to spread to more learners. Colleges and universities have become intermediaries in the process of providing higher education resources with learners as the main body. The application of educational big data and mobile Internet has broken the traditional "one-to-many" teaching mode of classroom that teachers and students "one-to-many" teaching mode. With the open service classroom form of "one-to-one" or even "many-to-one", the material entity of the classroom is gradually weakened and the virtual classroom is gradually born.

### **2.2 Data Mining and Analysis Technology Brings about the Individualization for Higher Education.**

At present, inefficiency of education funds and the ratio of education input-output contribute problems of backward teaching organization, such as teaching content, teaching method, examination system and talent training mode, which causes students' personality characteristics and cognitive development are ignored. Data mining and analysis technology bring personalization of higher education. Data mining technology can use learning behavior data and interactive data to provide learners with knowledge background analysis and learners' own characteristics recognition. It can also connect relevant knowledge points with learners, learners and knowledge points. It can make full use of high-quality resources through autonomous learning, active learning, collaborative learning, teaching consulting and so on. Personalized customized learning can be realized by potential.

### **2.3 Educational Big Data Promotes the Popularization for Higher Education.**

Many higher education teaching platforms, such as online courses, MOOCs, good universities and so on, have broadened the way of education. Qujing normal university has just completed the review and evaluation of undergraduate teaching in ordinary universities, and has just set up 10 courses, such as the public relations etiquette practice, and other 9 good college online courses such as the Tang poetry and Song Ci poetry, and so on as many as dozens of MOOC courses. The process data of these teaching can be realized during the class, which can do precise analysis service for our school teachers and students. The integration of big data technology, mobile Internet and Internet of Things technology will bring high degree of informatization for higher education. Virtual classroom, virtual classroom, network learning system and so on will realize higher education everywhere.

## **3. The Application of Educational Big Data Mining**

The application of various data mining technologies can effectively excavate the potential information of big data in education, eliminate the phenomenon of information isolated island, improve the quality of education and teaching, and promote the development of colleges and universities.

The research excavates large amount of data and obtains lots of potential information based on the undergraduate evaluation data of qujing normal university as well as the teaching status data of several secondary colleges.

### **3.1 Data Mining Technology.**

Data mining technology can be roughly divided into machine learning, database methods, set theory, bionics and statistical methods. Machine learning mainly includes support vector machines and inductive learning methods; database methods mainly include multidimensional data analysis,

OLAP, etc.; set theory mainly includes rough set theory and fuzzy set theory; biomimetic methods mainly include neural network, genetic algorithm, ant colony algorithm and so on; statistical methods mainly include regression analysis, correlation analysis, clustering analysis, discriminant analysis, factor score. Analysis, survival analysis and decision tree model.

### 3.2 Mathematical Statistical Method.

Mathematical statistical methods use probability theory to establish mathematical models, collect the data of the observed system, make quantitative analysis and summary, make inferences and predictions, and provide basis and reference for relevant decision-making. Relevant analysis, cluster analysis, discriminant analysis, factor analysis, survival analysis and regression analysis were used in this research.

#### 3.2.1 Relevance Analysis

Relevance analysis is used to study whether there is some dependence between phenomena, and to explore the relevant direction and degree of dependence between phenomena. It is a statistical method to study the correlation between random variables. The correlation can be determined by the measurement of indicators. The correlation coefficient is a statistical index used to illustrate the close degree of the correlation. There is only one correlation coefficient with positive and negative sign, which reflects positive and negative correlation. The correlation is expressed by R. It is assumed that there are two variables X and y. The correlation coefficient is calculated by product difference method based on sample data. The formula is as follows:

$$r = \frac{\frac{1}{n} \sum (x - \bar{x})(y - \bar{y})}{\sqrt{\frac{\sum (x - \bar{x})^2}{n}} \sqrt{\frac{\sum (y - \bar{y})^2}{n}}} \quad (1)$$

In the formula,  $\sum (x - \bar{x})(y - \bar{y})$  is the covariance of two variables, denominator is the standard deviation of two variables, which is presented by the formula below:

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (2)$$

By simplifying and expanding the above formula, the formula can be got as followed:

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}} \quad (3)$$

In the formula, the range of correlation coefficient r is  $0 \leq |r| \leq 1$ .

Two-sample correlation analysis is a statistical method to study the correlation between two variables. Pearson correlation coefficient is commonly used to calculate the data of fixed-distance variables, that is, to analyze the relationship between two continuous data. The formula for calculating the correlation coefficient is as follows:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

In the formula, the sign of correlation coefficient r reflects the direction of correlation, and its absolute value reflects the degree of closeness about correlation.

#### 3.2.2 Factor analysis

Factor analysis is a multivariate statistical analysis method, in which variables with intricate and complex relationships are reduced to a few comprehensive factors by studying the internal dependencies of variables. It describes each component of the original observation with the sum of the least measurable linear functions of common factors and special factors.

Factor analysis is described as followed:

$$X = U'_{(1)} F_{(1)} + U'_{(2)} F_{(2)} \quad (5)$$

$$X = (x_{1i}, x_{2i}, \dots, x_{ni})', i = 1, 2, 3, \dots, p; \quad (6)$$

Variable n is sample numbers and variable p is variable numbers.  $U'_{(1)}F_{(1)}$  is what main factors counted m can explains,  $U'_{(2)}F_{(2)}$  is it's residual part,  $U(1)$  is the factor load matrix and  $F(1)$  is the main factor. Thus the factor model can be obtained as followed:

$$\begin{cases} X_1 = \mu_{11}F_1 + \mu_{12}F_2 + \cdots + \mu_{1m}F_m \\ X_2 = \mu_{21}F_1 + \mu_{22}F_2 + \cdots + \mu_{2m}F_m \\ \vdots \\ X_p = \mu_{p1}F_1 + \mu_{p2}F_2 + \cdots + \mu_{pm}F_m \end{cases} \quad (7)$$

In the formula,  $F = (X_1, X_2, \dots, X_p)'$  and  $X$  is a common factor,  $\mu_{ij}$  the factor load matrix, which is the load of the first variable on the  $j$  common factor. Residual is a special factor, which is independent of each other, and obeys normal distribution  $N(0, \sigma_i^2)$ .

### 3.3 Data Mining Case.

Taking the big data of students 'employment as an example, the data comes from the questionnaire of graduates' employment and training quality of qujing normal university, which included more than 3100 valid questionnaires. Each college's annual report of employment quality in 2017. For the missing data, the python crawler program based on machine learning is used to automatically grab the relevant information of school office system, teaching data and service data, and carry out data preview. After treatment, factor analysis in mathematical statistical method is used to analyze the results, as shown below.

Table 1 Analysis of Employment Data Mining In Local Universities- Communalities

Communalities		
	<i>Initial</i>	<i>Capture</i>
CET4(1)	1.000	.905
Investment in employment funds(2)	1.000	.861
Student source(3)	1.000	.882
Provincial –level awards(4)	1.000	.855
Initial employment rate(5)	1.000	.817
Graduates' Job Satisfaction with Employment Services(6)	1.000	.872
Help Number for Students with Difficulties(7)	1.000	.861
PSC Test(8)	1.000	.854
Acquisition of Vocational Skills Certificate(9)	1.000	.661
Graduates' Satisfaction with School Talents Training(10)	1.000	.791
Employment guidance and consultation(11)	1.000	.790

Table 2 Analysis of Employment Data Mining In Local Universities- Initial Eigenvalue

Component	Initial eigenvalue		
	<i>Total</i>	<i>Variant %</i>	<i>Accumulation %</i>
1	3.019	27.443	27.443
2	2.188	19.892	47.335
3	1.603	14.576	61.911
4	1.294	11.767	73.678
5	1.044	9.489	83.168
6	.797	7.246	90.413
7	.471	4.277	94.691
8	.293	2.665	97.356
9	.192	1.747	99.104
10	.091	.829	99.932
11	.007	.068	100.000

It can be seen that the common factor contributes 90% to the variance of CET4, the characteristic root of the first principal component is 3.019, the cumulative contribution rate of the first five factors is 83.168%, and the eigenvalue is greater than 1, so the first five factors are selected. Non-foreign language college students pass CET-4 and statistics of CET-4 (Tai Language) are the most helpful to students' employment, which can effectively promote the quality of students' employment; secondly, the proportion of college investment in employment funds, the amount of investment also affects the final employment rate of the college, and departments with relatively large investment funds. The employment situation of its graduates is better; the second is the source situation; the non-Yunnan students have more active thinking, unique personality, usually better comprehensive performance and easier employment; the second is that the students get awards at the national level and provincial level, which has a certain impact on the employment of students; the last is the first employment rate, which directly affects the overall employment rate of the school. Graduates' job satisfaction and other factors also have a certain impact on students' employment situation, but they are not the main component. Schools should pay special attention to the first five factors and adopt effective policies to significantly promote students' employment rate and employment quality.

#### **4. Conclusions**

Colleges and teaching assistant departments in Colleges and universities have a large amount of data on running status. Through data mining technology, potential value of mass information can be excavated, precise services can be provided, and scientific decision-making can be supported. Through data mining, analysis, machine learning and other technologies to show the value of big data in education, promote the development and innovation of education, improve the quality of education and teaching, achieve wisdom teaching and learning, wisdom service and wisdom evaluation.

#### **Acknowledgment**

It is supported by Philosophy and Social Science Educational and Scientific Planning Office of Yunnan Province of China (Application Research on Educational Big Data to Influence Management Decision Making of Higher Education, Grant No. AD17012), and is also supported by qujing normal university Project (Research on the Way of Integrating Socialist Core Value System into Ideological and Political Education in Universities, Grant No. 2011MS006).

#### **References**

- [1] Information on <http://www.Gartner.com/it—glossary/big—data>.
- [2] Y.N. Zhang: A Study of Big Data Applications In the Field of Education Based on the Practice in America (Ph.D., East China Normal University, China), pp.28–32 , 2016.
- [3] Y.G. Wang, Q. Zhang: MOOC: Characteristics and learning mechanism, Education Research, No9, pp.112-118,(in Chinese), 2014
- [4] Y.X. Lv, S.C. Peng, Y. Yuan: A Classifier Using Online Bagging Ensemble Method for Big Data Stream Learning, Tsinghua Science and Technology, Vol.24, No.4, pp. 379-384, 2019.
- [5] U. Shefa, J.Y. Jung: Comparative study of microarray and experimental data on Schwann cells in peripheral nerve degeneration and regeneration: big data analysis, Neural Regeneration Research, Vol.14, No.6, pp. 1099-1103, 2019
- [6] T.M. Wang, Y. Tao, H. Liu: Current Researches and Future Development Trend of Intelligent Robot: A Review, International Journal of Automation and Computing, Vol.15, No.5, pp.525-529, 2018.
- [7] Information on <http://www.slideshare.net/BemardMarr/a-brief-history-of-big-data>.

- [8] H. MEI, L. ZHANG: Can big data bring a breakthrough for software automation? Science China (Information Sciences), Vol.61 , No.5, pp.288-293, 2018.
- [9] H. Hakeem. Layered Software Patterns for Data Analysis in Big Data Environment, International Journal of Automation and Computing, Vol.14, No.6, pp.650-658, 2017.
- [10] J. Zhu, J.F. Chen, W.B. Hu: Big Learning with Bayesian methods, National Science Review, Vol.4, No.4, pp.627-632, 2017