Design of University Evaluation System based on Big Data

Ji Wu

School of Computer Science and Network Engineering, Guangzhou University, Guangzhou, Guangdong, 510006, China
wuji@gzhu.edu.cn

Abstract. In the practice of early university evaluation, due to the backwardness of information technology and evaluation methods, the following problems were brought about: the social needs were neglected, the evaluation subject was single, the evaluation method was simple, and the evaluation data source was single. In the paper a big data model of university evaluation is presented, which studies the educational activities from 8 dimensions: core performance, innovation ability, teamwork, resources, audience willingness, parental evaluation, employer evaluation and social contribution. The model takes account of long-term, short-term factors, individual and overall factors, internal and external factors, educators and audience factors, etc., and carries out omnidirectional analysis. Every dimension is analyzed from 2 aspects: positive and negative. Then based on the model, an University Evaluation System based on Big Data (UESBD) was designed and implemented. Finally, the software was used to analyze the data of 20 colleges and universities from multiple dimensions to evaluate the educational activities of education.

1. Introduction
In the practice of early university evaluation, due to the backwardness of information technology, the evaluation methods were limited, which brought about the following problems:

First, the establishment of the evaluation index system has neglected social needs to some extent. The university evaluation should strive to provide practical reference information for the society. Therefore, the starting point of the evaluation activity should come from the society, and the establishment of the indicator system should also pursue the main purpose of meeting the needs of the society. However, most of the evaluation index systems used in the comprehensive evaluation activities of universities in China were based on the understanding of the quality of universities.

Second, the evaluation subject was single. The factors directly related to education, such as students, parents, and employers were ignored. Only experts were used to evaluate the level of university education.

Third, most of the evaluation methods used simple quantitative scoring, then linear weighting, and finally sorting different universities.

Fourth, the source of evaluation data was single and the amount of information was greatly limited. The school to be evaluated prepared a large amount of information in advance according to the requirements of the evaluation indicators. These data could not truly and comprehensively reflect the educational situation of the school.

Big data technology can handle not only structured data, such as traditional information system and questionnaires, but also semi-structured and unstructured data, such as social network information, geographic information, audio and video information, etc. The previously neglected indicators of social needs can also be solved. Commonly used social needs evaluation indicators include:
evaluation of employers, salary of middle and long-term graduates, situation of graduates in public welfare, evaluation of graduates by society, satisfaction of graduates in graduate schools in the medium and long term, graduates' satisfaction with their lives, etc. [5-8]. The evaluation of these problems is almost impossible to achieve under the traditional technical conditions, while in the big data environment, through the sharing of various information system information and the mining and analysis of semi-structured unstructured data, it is possible to collect and analyze.

With the continuous development of education development and reform in China, how to scientifically and effectively evaluate the quality of education and establish an efficient university evaluation system have become one of the major tasks that need to be solved urgently in education reform. In the current environment, the construction of a university evaluation system based on big data (UESBD) has strong theoretical value and practical significance.

2. Big data model of University Evaluation System

This paper has initially established a set of big data models to study the effects of educational activities from the core dimensions, innovation ability, team, resources, brand influence, parent evaluation, employer evaluation and social contribution. The model analyzes and establishes data collection points from long-term and short-term factors, individual and overall factors, internal and external factors, educators and audience factors. Each dimension is analyzed from the positive and negative aspects, so that a total of 16 observation points can be established, and each observation point collects corresponding sample data. The data comes from all aspects of integrated life, including the entire big data on the Internet, as well as research data.

2.1. The first dimension: the direct effect of education: core performance
The core effect is the most important factor and the main evaluation method in the current education evaluation method, and it is also the short-term effect of educational activities. This indicator is for the group who receive education. Positive indicators include employment rate, average salary, student awards, and quality of dissertations. Negative indicators include the number of active dropouts, negative impact events, unqualified exams, and the number of graduates in time.

2.2. The second dimension: innovation ability
This indicator is for organizations that implement education. Positive indicators include major achievements, representative results, participation in standard norm formulation and recommendations. Negative indicators include paper plagiarism.

2.3. The third dimension: teamwork
The positive indicators include the number of academicians, the number of national scholars/talents, the number of provincial scholars/talents, international academic journals, the number of key disciplines, and the number of key laboratories. Negative indicators include the proportion of teachers in the school, and the number of non-teachers lost in the past three years.

2.4. The fourth dimension: educational resource coordination capacity: resources
This indicator focuses on the hardware resources of educational activities and the level of software management as an aid to educational activities. This indicator is for the organization that implements education. Positive indicators include rationality of institutional establishment, logistics support (accommodation, diet), basic hardware resources, etc. Negative indicators include the bureaucratic bureaucracy, the cumbersome process, and complaints from students.

2.5. The fifth dimension, education brand recognition: Audience willingness
This indicator is for the group receiving education. Positive indicators include initially from the following aspects: the active awareness in the education process: target and participation; the degree of satisfaction in the education process: the feedback of factors affecting education, such as educators,
educational environment (soft environment, hard environment), participating partners, etc.; self-learning ability: the initiative of continuous active learning, the cultivation of good behavior habits; the alumni donation. Negative indicators include influential group events, etc.

This indicator is subjective and can be conducted using a questionnaire or commissioned by a third party.

2.6. The sixth dimension, direct education assessment: parental evaluation
This indicator is for parents of students who pay for educational services. Positive indicators include children's attitudes towards family and neighbors, changes in language skills, ability to handle affairs, changes in psychological relationships, etc.; schools provide online communication platforms for parents, provide information on student performance, provide information about school activities, provide parental counseling, invite parents to participate in school activities, etc. Negative indicators include parent complaints, etc.

This part of the indicators are subjective evaluations and can be conducted using questionnaires or commissioned by third parties.

2.7. The seventh dimension, indirect evaluation of education: employer evaluation
Positive indicators include the depth and breadth of professional knowledge: interdisciplinary knowledge, can meet the needs of work or not; professional skills: the number and proficiency of professional skills, can meet the needs of work or not; professionalism: professional attitude, professional awareness, professional habits, professional conduct, etc. Negative indicators include dismissed or expelled by the employer.

This indicator makes full use of the big data platform to search for relevant data directly from the Internet.

2.7.1. The eighth dimension, the long-term effect of education: social contribution
Educational effectiveness is the short-term effect of education and can reflect the results of educational behavior in a timely manner, but it cannot measure long-term effects. Social contribution is the fundamental goal of education. Education itself is to cultivate excellent citizens for the society. The target of this indicator is the group receiving education. The positive indicators include the list of high-confidence social rankings, the number of outstanding alumni (leading talents, key talents), and the annual contribution of alumni to the school. Negative indicators include adverse public opinion events.

3. Implementation of University Evaluation System Based on Big Data
On the basis of theoretical research, the research team built a set of automated application systems based on the mature IT infrastructure platform - a university evaluation system based on big data (UESBD). It mainly includes three parts: data collection and storage, business model establishment and data analysis, and data mining and query.

The task of the data collection and storage subsystem is to collect structured data (such as information system basic data, employer data, third-party evaluation results data, etc.), unstructured data (such as social network data, web data, audio, video data, etc.) and external input data (such as questionnaire results data, communication results feedback data). Then it establishes a unified data storage model, collects data in categories, and finally sorts it into the database. The basic tool is MySQL.

The business model building and data analysis subsystems are used for data mining through specialized analysis methods. The tools are basic tools MySQL, Hadoop, SparkSQL and R;

The data mining and query subsystem is based on data analysis and conducts educational evaluation and decision making. The data mining tool is R. The data query tool is ElasticSearch. The results of the analysis are presented to the competent authorities, universities and the society, which can assist the competent departments in data decision support, resource allocation and policy guidance,
help the university peers to benchmark and complement each other, help students to choose schools, help employers to select talents, and assist social supervision.

According to the constructed evaluation model, the collected big data is cleaned and analyzed. The scores of each assessment dimension are calculated by eight dimensions, and then calculated separately to obtain data of eight dimensions. This data is based primarily on 2018 data. Some indicators of some observation objects lack data, and are directly replaced by zero values or other year values; some data indicators are grading values, which are not quantified, and are also quantified according to the level.

From the evaluation data, 10 universities are randomly selected, and a stacked bar chart is generated according to the collected positive indicator values of the 8 dimensions. Figure 1 can very intuitively show the size of each university's indicators; if the cumulative total system is used, the total length is the cumulative total score.

![Figure 1. Stacked bar chart of positive indicators of ten universities.](image)

For the same object observation, the radar chart can more intuitively show whether its aspects are balanced, which aspects are advantages, and which areas need improvement. As shown in Figure 2, radar chart of indicators of Donghua University (DHU), it can be seen that the core performance are more prominent, parental evaluation and employer evaluation is a short board, and other aspects are relatively balanced. In terms of the negative list, the audience willingness (brand influence) and the employer evaluation are also injurious.

In summary, after establishing rich source data, UESBD can faithfully reflect the real situation and level of each university, which can reflect its advantages and its shortcomings.
4. conclusion
With the rapid development of school informatization and digital campus, the environment of big data gradually formed a large scale. In the paper UESBD is presented, which creates university evaluation model and studies the educational activities from the 8 dimensions of core performance, innovation ability, teamwork, resources, audience willingness, parental evaluation, employer evaluation and social contribution.

Like all big data systems, software is easy to implement, but data are hard to find. The shortcomings faced by this system are also the same, so this paper only analyzes 20 universities that can collect more comprehensive data. The next step is to develop interfaces with various systems, such as recruitment websites, employer systems, and university teacher systems, to obtain data directly from various information systems, thereby enriching the data sources of the university evaluation system. Only when the amount of data is increasing, can we get more and more effective evaluation results.

Acknowledgments
This work is supported by Guangzhou Education Science Planning Project: Design and Application of High-level University Evaluation System Based on Big Data under Grant No. 1201534138.

References
[1] Fan, X., Zhang, M. (2012) Analysis of the Differences in the Comprehensive Evaluation Index System of Chinese University Rankings. Chinese Adult Education, 7:28-33.
[2] Shi, Y., Yuan, Y., Song G. (2017) Based on arwu's world university ranking system comparison and empirical research. Library and Information Work, 15: 95-103.
[3] Li, Y., Chen, K., Wu, Y., Zheng, W. (2015) Big Data Processing Model - System Structure, Methods and Trends. Small Microcomputer System, 36:641-647.
[4] Zhan, J., Gao, Y., Wang L., Li, J., Wei, K., Luo, C. (2016) Bigdatabench: Open source big data system benchmark. Computer Journal, 1:196-211.
[5] Li, H., Liu, B. (2013) "US News and World Report" research on the evaluation index system of undergraduate college rankings. China Higher Education Research, 11: 57-62.
[6] Ji, S., Li, W. (2014) Evolution of the University Rankings of Higher Education in the UK. Education Review, 7: 150-152.
[7] The results of the 2012 Discipline Assessment of the Ministry of Education's Degree and Graduate Education Development Center. http://www.chinadegrees.cn/xwyyjsjyxx/xxsbdxz/index.shtml#3.
[8] Higher Education Teaching Evaluation Center of the Ministry of Education: Eligibility Evaluation Index System for Undergraduate Teaching Work in Ordinary Colleges and Universities. http://www.pgzx.edu.cn.