Design of Data Analytics Teaching Platform for IT Talents Driven by Enterprise Projects

Ruijun Zhang[1(✉)] and Wenxia Li[2]

[1] Center of Service Science and Engineering, Wuhan University of Science and Technology, Wuhan, China
zrjun@wust.edu.cn

[2] College of Railway Communication and Signaling, Wuhan Railway Vocational College of Technology, Wuhan, China

Abstract. The research investigates the data analytics capability of it related students in Wuhan University of science and technology in the form of questionnaire. The data show that the theoretical and practical teaching system of data analytics courses has been basically established. However, there are many but scattered knowledge points and lack of organic connection of cases. The whole life cycle based data analytics cases, such as the environment construction of big data, data preprocessing, storage and transmission, data mining, machine learning and data presentation, are difficult to be carried out in the laboratories of most colleges in China. There is a serious shortage of teachers with enterprise project experience. Based on this, a data analytics teaching platform driven by enterprise projects is proposed in this paper. Four aspects are involved in the platform: college and enterprise jointly develop data analysis projects; building data analysis platforms shared with data sets and project cases; adopting the online and offline mixed teaching mode; establishing a multi-layer progressive mode of curriculum design, subject competition and project practice to cultivate students' engineering practice ability from the whole life cycle of data analysis projects. After online learning about key technologies and business, with the help of offline technology exchange and algorithm design, the project group completed the feasibility analysis, design, programming and implementation of the project based on whole life cycle, which achieved outstanding results in the project fields of analytics the public opinion of COVID-19, identification of innovation investment opportunity, and so on.

Keywords: Data analytics · Project driven · Teaching platform · Questionnaire investigation

1 Introduction

The widespread use of big data technology in various industries has resulted in a shortage of practical personnel on big data application. In 2019, the employment gap of big data related jobs has reached the number of 1.5 million in China. As an ability on data mining and data insight, data analysis provides a brand new approach to promote technological

Z. Zhan et al. (Eds.): ICBDIE 2022, AHCS 5, pp. 314–322, 2023.
https://doi.org/10.2991/978-94-6463-034-3_32

development, creates business value and optimize social services. Data analytics capability (DAC) is playing an irreplaceable role in many fields, such as enterprise marketing [1], public opinion analysis [2], enterprise sustainable development [3], supply chain management [4], business innovation and processes improvement [5], enhancing firm performance [6], workflow customization [7], higher education [8], etc. Big data analysts have become the core competitiveness and the top-level talents of Internet enterprises (such as Google, Baidu, Alibaba, etc.).

There are a few research papers on the cultivation of DAC. [9] designed an analysis and application systems of campus big data in different themes and levels, which was developed for educational policy proposals. An exploratory teaching program about big data analysis was proposed at Kocaeli University in the spring semester of 2018–2019, which was designed and given in Department of Computer Engineering [10]. [11] proposed a Π teaching model of big data analysis. Few scholars focused on the cultivation mode of big data analysts driven by enterprise needs, as well as the way to improve the practical engineering capabilities of students.

2 Questionnaire Design

In order to relieve the "two separate layers" dilemma between the demand of enterprises and the supply of colleges and understand the training situation of DAC of IT specialties, the research group designed an online questionnaire to obtain the primary data of colleges' teaching and social practicing modes training the big data analysis abilities of student and relative cultivation systems.

2.1 Content Design and Respondent

To ensure the validity of this questionnaire, the respondents were set as senior and junior college students who had learned about data analysis, including majors of Information Management and Information System, Computer Science and Technology, E-commerce and other IT specialties in WUST. Questionnaire Star software was used for giving out and collecting the online questionnaire. The research group designed 40 questions, gave out 130 questionnaires and collected 124 of them, at a rate of 95.38%. Among the 120 valid samples, 55% were boys and 45% were girls. 52.5% were senior students and 47.5% were junior students.

2.2 SOLO Classificational Capability Evaluation Method

In order to quantitatively evaluate the sub-items of DAC, SOLO(Structure of the Observed Learning Outcome) classificational capability evaluation method is used to assess a certain ability with five levels from low to high: Prestructural, Unistructural, Multistructural, Relational and Extended Abstract (Lucander 2010). For each level, different calculation weight is assigned, as shown in Table 1.

Table 1. Analysis method of SOLO and Calculation Weight

Level	Description	Weight
Prestructural	Unable to understand or solve problems and only provide answers without logic and arguments.	1
Unistructural	Find only one idea to approach the solution and stop looking for others, urgent to conclude an answer.	2
Multistructural	Find a number of ideas to approach the solution but unable to integrate them logically.	3
Relational	Find a number of ideas to approach the solution and able to integrate them logically to obtain well-formed ideas.	4
Extended Abstract	Able to summarize the problem abstractly, analyze the problem theoretically, and expand the connotation of the problem.	5

Table 2. Self-commenting Score of DAC Factors

Factor	Average Score	Average Score (male)	Average Score (female)
Data recognition	3.05	3.16	2.93
Data collection	2.90	3.00	2.79
Data evaluation	2.88	2.94	2.83
Data preprocessing	2.80	3.10	2.48
Model training & verification	2.68	2.84	2.52
Mathematical modeling	2.35	2.71	1.97

2.3 Reliability Test

A reliability test was carried out on the 40-question questionnaire after excluding sequencing questions and questions allowing more than one choices. The Cronbach α coefficient is 0.725, which is larger than 0.7. This result shows that the research data is reliable enough.

3 Result Analysis

The questionnaire is generally launched from four aspects: the current situation about students' DAC, curriculum teaching, experiment & practice and training mode. Due to the limitation of article length, only some representative questions are listed here.

3.1 The Current Situation About Students' DAC

On the basis of SOLO, the research group organized the respondents self-commenting on some factors about DAC. Score-based comprehensive evaluation method is used in the

analysis: Comprehensive average score = (Σ frequency * calculation weight)/ number of samples. As is shown in Table 2, the students' ability of data recognition is just fair, and their abilities of data collection, data evaluation and data preprocessing are not good, and their abilities of model training & verification and mathematical modeling are weak and need to be improved. What's more, it is found that the relative analysis abilities of male are better than those of female.

3.2 Curriculum Teaching

The research group carried out the same analysis procedure as in Sect. 3.1 for the teaching organization mode of big data analysis courses. The result is listed in descending order of scores: Driven by projects, combining theoretical knowledge with programming code (3.15 points); led by theoretical knowledge in class, supplemented programming practice after class (2.73 points); led by students' self-study about programming, supplemented by teachers' guidance (2.10 points); led by online live classes, supplemented by recorded class from online platform (2.02 points). Through the interviews, it is learned that the full life cycle course teaching mode driven by the actual project of the enterprise is very popular among students.

As for the comparison between online teaching and classical class teaching, 35% students believe that online teaching is more efficient and convenient for learning; 33.33% students deems that the online teaching is of low efficiency and they can learn little from online classes; 25% students deem there is no difference; 6.67% of them are not sensitive to the difference. Compared with junior students, senior students' options towards online teaching efficiency are more polarized: some think that teaching efficiency is high, but others don't feel it.

3.3 Experiment & Practice

Data analysis is a course with high practicality. Specifically, the experiment & practice part of the course includes computer operation, course project, engineering project participation, etc. 42% students thought that these sessions are very helpful to improve data analysis ability; 52% students indicated that they only help a little; and only 4% thought that they provide little or no help.

When asked about the topic - How the experiment & practice part improves the big data analytics capability (DAC)?, the students voted on the importance of the three aspects of experiment & practice. The result is listed in descending order of scores: engineering project participation (3.72 points), course project (2.73 points) and in-class computer operation (2.10 points). It can be inferred that students are very eager to participate in real big data analysis projects from enterprises.

As for the question- What intractable or unsolvable problems have you encountered during the experiment course of big data analysis? (multiple choice) - 24.3% students thought they meet too many errors and bugs when coding and debugging; 19.7% students held that teacher provide more samples at class but less practice chance; 18.2% students fell into the trouble of parameter debugging (such as learning rate of BP neural network, etc.); 17.5% students considered it difficult to understand the source code.

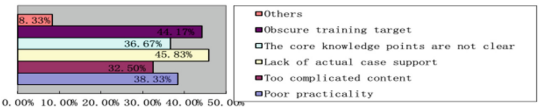

Fig. 1. Problems and defects in the teaching of big data analysis courses

Among the investigated students, junior students usually encounter more problems than senior students at the class of big data analysis.

3.4 Cultivation Mode

For the cultivation of DAC, different majors have set up different curriculum systems and formulated corresponding training schemes. The research group designed a questionnaire on the problems and defects of cultivation mode, as shown in Fig. 1. The results show that the lack of practical case support makes the knowledge points unable to be connected and the training objectives vague, which has become a pain point.

As for the question - What methods do you think can help you improve your DAC? (multiple choice)- 60.83% students thought that teacher should link up related knowledge points through specific projects; 56.67% students thought that some common ways to solve problems should be given by teacher; 52.50% students thought some data analysis competitions should be held by the college; 44.17% students hope to increase the programing opportunities after class. Most of them considered that the training mode driven by enterprise projects and competitions provides a good solution for improving college students' DAC.

As for the question - Is the big data analysis knowledge you have learned helpful during your internship or job seeking?"- very few students (3.33%) thought that knowledge about data analysis was very helpful; 61.67% students thought it was helpful; 16.67% students thought it was less helpful; very few students (6.67%) thought it was not helpful at all.

3.5 Result Analysis

According to the questionnaire survey above and the information obtained by interviews with college students, teachers and entrepreneurs, the following problems about cultivation mode of DAC in colleges are found:

(1) Teaching resources (such as big data sets, project sets) based on enterprise projects are not sufficient.
(2) Full-chain practical teaching system has not been well established. Some life cycle based big data cases, such as the environment construction of big data, data preprocessing, storage and transmission, data mining, data analysis and data presentation, are difficult to be carried out in the laboratories of most colleges in China.
(3) Students have a strong willingness for project-based teaching driven by practical cases of enterprises.

4 A Four-Dimensional Integrated Teaching Platform of DAC

In order to improve the DAC of IT major students, a teaching platform driven by big data analysis projects is built based on four dimensions of < project, platform, teaching, practice >. The system relies on the school enterprise cooperation platform, adopts the combining mode of online teaching and offline experiment, and cultivates students' engineering practice ability from course project to graduation design, from suject competition to project practice, as shown in Fig. 2.

4.1 Establishment of College-Enterprise Cooperation Big Data Platform

Enterprises have advantages in projects and hardware, and colleges have advantages in intelligence and human resources. These advantages provide a good basis for their cooperation. College have high-performance computing centers, mature training programs such as curriculum syllabus, textbooks, exercises, and experiments, strong scientific research capabilities and technical reserves. IT enterprises have stable and robust cloud servers, enough industry big data and various engineering project solutions. Colleges can rely on big data analysis laboratories to cooperate with IT enterprises. They can build a big data cooperation platform through PaddlePaddle, MOOC, QQ, WeChat and other channels. They can also share software, hardware and data resources, and build a case base and project base oriented to actual needs.

4.2 Big Data Analytics Projects Oriented to Practical Needs

Project teaching is driven by real-life problems and guided by theory. It helps to cultivate students' ability to find problems, analyze problems and solve problems in real life. The

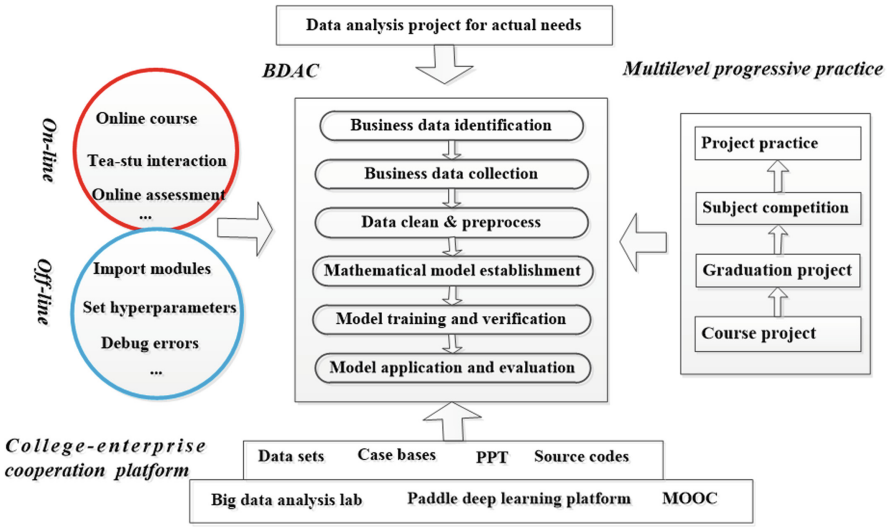

Fig. 2. A four-dimensional integrated teaching platform of DAC

bottleneck of project teaching is lack of project source. Taking big data analytics project as an example, it is a problem to build a rich dataset, project library and case library. Generally speaking, the shortage of dataset and project source can be solved through the following three ways:

1) Establishing the classical case bases through teacher training and conference communication etc. For example, Boston house price prediction, iris classification, cat and dog recognition, handwritten digit recognition and other cases.
2) Establish the unique engineering project bases through the cooperation of college-enterprise horizontal projects. For example, the problem of motor vehicle registration in parking big data, the problem of automatic identification of electricity theft and leakage users in industrial electricity consumption, and the problem of fugitive identification in the public security big data center.
3) Establishing the innovative project bases through students' extracurricular science and technology activities. For example, the problem of the evolution of public opinion on the COVID-19, the problem of identifying the information of online loans for college students, etc.

4.3 Combination of Online Learning and Offline Experiment

To improve teaching effect, an O2O mixed teaching mode is adopted. Online course platforms like MOOC, Zhihuishu and Rain Classroom are used. Teachers give classes on convolution neural network, recurrent neural network, support vector machine, recommendation algorithm, clustering and classification and other knowledge points with the form of micro-lecture. And also, students can interact with teachers on those platforms. After class, students set up project groups, install Anaconda integrated environment, and load data analysis and calculation packages such as Tensorflow, Sklearn, Numpy, Keras and Tarfile according to the required versions. Then, under the guidance of teachers, students design models, write code and debug in Jupyter Notebook webpage environment.

4.4 The Multi-level Progressive Practical Teaching Mode of Course Project, Competition and Project

As a comprehensive ability, the construction of knowledge structure of big data analytics capability needs the support of professional courses in multiple fields. Taking the students majoring in information management and information system as an example, in the formulation of the curriculum system, many professional courses should be set up successively, such as C language, database principle, management statistics, data structure, python programming, data acquisition and processing technology, Fundamentals of big data technology, data analysis and so on. In practice, the knowledge points of the course can be organically connected through course project. Students' algorithm design ability and programming ability can be improved by organizing students to participate in various data analysis competitions (such as TipDM Cup Data Mining Challenge competition and Alibaba Tianchi big data competition, etc.). Students' team spirit, communication ability, analysis and problem-solving ability can be cultivated managed by teacher through the enterprise entrusted project.

4.5 Cultivation of Students' Data Analysis Ability Driven by Complete Projects

In order to cultivate students' practical engineering ability, complete big data analysis projects (such as vehicle license plate recognition in parking lots, drug recommendation based on historical medical records of patients, etc.) are carried out to cultivate students. The workflow of these projects includes selective extraction of big data of business systems, exploratory data analysis, data preprocessing, system modeling, model training & verification, model application & evaluation, etc.

For example, to analyze the public opinion of COVID-19 in the first half of 2020, a college- enterprise cooperation project team was established, which was composed of teachers and students of WUST and technicians of Baidu Netcom Science and Technology Co., Ltd. The project was supported by the school enterprise cooperation and collaborative education project of the Ministry of education. In this project, an elder-lead-younger organization was established: teachers lead students, postgraduates lead undergraduates and senior students lead junior students. Over a period of three months, the group obtained tens of thousands of microblog texts and more than 100,000 comments with web crawlers from official websites like People's Daily Online, CCTV and Xinhua Net. After attaching positive and negative marks to 11871 comments, the group adopted the methods of natural language processing to complete a full life cycle machine learning project in the fields of emotion tendency analysis of microblog comments, guidance of social public opinion. The whole workflow includes preprocessing comment text, word segmentation, modeling, training, verification and prediction. There were a total of 4 teachers, 6 postgraduates and more than 40 undergraduates participating in the project, from the beginning to the end. Through the full participation of the project, teachers and students have an overall understanding of big data analysis, and their big data 他the project, many abilities, such as information retrieval, web crawling, data cleaning and marking, modeling and simulation, prediction, evaluation, and so on, have been greatly improved.

5 Conclusion

Data analysis courses are usually of strong practicality and include a large number of knowledge points. Traditional mode of course teaching can hardly match the demand of cultivating compound talents. The new project-driven training mode, which combine online teaching and after-class practice, is an attempt on the cultivation of students' big data abilities. Since the mode was proposed, remarkable achievements have been reached.

Acknowledgements. This research is supported by High School Education Research Program, under Grant 2020369 from Hubei Province, and by the Key Research Projects of Philosophy and Social Sciences in Colleges and Universities from the Hubei Provincial Department of Education, under the Grant 21D014, China.

References

1. Shahbaz, M., Gao, C., Zhai, L., Shahzad, F., Luqman, A., & Zahid, R. (2021). Impact of big data analytics on sales performance in pharmaceutical organizations: The role of customer relationship management capabilities. Plos One, 16(4).
2. Ko, J., Paek, S., Park, S., & Park, J. (2021). A News Big Data Analysis of Issues in Higher Education in Korea amid the COVID-19 Pandemic. Sustainability, 13(13).
3. Zhang, H., Song, M., & He, H. (2020). Achieving the Success of Sustainability Development Projects through Big Data Analytics and Artificial Intelligence Capability. Sustainability, 12(3).
4. Arunachalam, D., Kumar, N., & Kawalek, J. P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. Transportation Research Part E-logistics and Transportation Review, 114, 416-436.
5. Lozada, N., Arias-Perez, J., & Perdomo-Charry, G. (2019). Big data analytics capability and co-innovation: An empirical study. Heliyon, 5(10).
6. Gu, V. C., Zhou, B., Cao, Q., & Adams, J. (2021). Exploring the relationship between supplier development, big data analytics capability, and firm performance. Annals of Operations Research, 302(1), 151-172.
7. Chok, K., Tomek, P., & Clifton, T. (2017). Cloud accelerated Performance based Seismic Design. Paper presented at the 39th IABSE Symposium in Vancouver 2017: Engineering the Future, September 21, 2017 - September 23, 2017, Vancouver, BC, Canada.
8. Cui, L., Peng, Y., Ding, L., & Lu, D. (2021). An improved batch fluidized drying experimental design based on digital sensors and a minicomputer. Engineering Reports, 3(8).
9. Zhang, W., & Jiang, L. (2018). Algorithm Analysis for Big Data in Education Based on Depth Learning. Wiress Personal communications, 102(4), 3111-3119.
10. Schultze, J. L. (2015). Teaching 'big data' analysis to young immunologists. Nature Immunology, 16(9), 902-905.
11. Wang, L., & Lu Q. (2021). Exploration of Π type teaching mode of Big Data Analysis course under the background of new liberal arts. Jiangsu Science & Technology Information, 38(16), 71-73.