Exploration and Practice of Big Data System and Technology Blended Teaching Mode

Qiuyun Zhao(✉) and Le Wei

School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China
zqyuen@cuit.edu.cn

Abstract. Big Data System and Technology is a professional course for the data science, Big data technology and related majors. It has the characteristics of complex content and high practical difficulty. There are several problems in current teaching, such as outdated curriculum concept, single teaching method, disconnection between theory and practice, and difficulty in achieving curriculum objectives. To solve these problems, blended teaching is introduced into the curriculum teaching. The course content of Big Data System and Technology under the blended teaching mode is designed. The division of online and offline teaching content is defined. How to practice the blended teaching mode is determined before, during and after class, and the course assessment and evaluation method focusing on process assessment is given.

Keywords: Blended Teaching · Big Data · Online Teaching · Case-based Teaching · Problem Oriented

1 Introduction

Entering the era of big data, the big data industry has become an important engine to accelerate the economic and social development, especially in terms of the quality, efficiency and power transformation. The industry development cannot be separated from the support of talents. At present, the market is in urgent need of big data architects, big data development engineers, big data algorithm engineers, big data analysis engineers and big data operation and maintenance engineers. In order to meet the needs of the society, many Chinese universities have set up data science and big data related majors, and carried out exploration and practice in the aspects of the talent training target, talent training program, curriculum system, teaching reform and teaching methods, and gained certain results and experience.

Big Data System and Technology is a basic course of big data that should be offered by many colleges and universities. This course takes the current popular system of Hadoop as the object of explanation, focusing on the basic architecture of Hadoop, distributed file system of HDFS, distributed computing framework of MapReduce, Hadoop framework ecosystem, the Hadoop practice and other contents. This course helps students master the knowledge of big data distributed storage, distributed processing, big

L. F. Ying et al. (Eds.): ICELA 2022, ASSEHR 730, pp. 120–129, 2023.
https://doi.org/10.2991/978-2-38476-004-6_17

data analysis and big data application, which lay a solid foundation for further study. For undergraduates, Big Data System and Technology, as a relatively new course, has the characteristics of the complex and complicated theoretical system, high practical difficulty, and the big data thinking throughout. Therefore, it is necessary to carry out relevant research and practice from aspects of curriculum objectives, curriculum content, curriculum resources, teaching methods, teaching modes, teaching means and so on. This is of great significance for the implementation of training scheme and the achievement of curriculum objectives.

The blended teaching is a new teaching mode, which uses different teaching theories, different teaching techniques and teaching methods to implement teaching. Its essence lies in rethinking and redesigning the relationship between teaching and learning [1]. Professor He Kekang of Beijing Normal University believes that the blended teaching mode combines the advantages of traditional teaching methods with those of networked teaching. Blended teaching not only gives full play to the leading role of teachers in guiding, inspiring and monitoring the teaching process, but also fully reflects the initiative, enthusiasm and creativity of students as the subject of the learning process [2]. After years of practice, people's cognition of blended teaching has gone through three stages of the technology-centered stage, teacher-centered stage and student-centered one [3]. At present, blended teaching is widely regarded in China as a teaching model of 'traditional classroom + online learning + flipped classroom', with more emphasis on student-centered teaching. Although there are different definitions of blended teaching, these definitions all contain two key connotations. Firstly, it is based on the constructivism learning theory, emphasizing that students' learning is the center, and the acquisition of knowledge and ability is completed in the process of students' active construction. Secondly learning is the integration of online learning and offline learning, and marked by the application of information technology [4].

With the rise of MOOC in recent years, blended teaching has been widely used in China. When it comes to big data courses, Luo Guotao [5] et al. carried out blended teaching of Big Data Processing, made full use of online teaching platform and MOOCs to realize the combination of online and offline, in-class and out-of-class, pre-class, in-class and after-class. Zhang Lingli [6] et al. explored the content design and teaching methods of general knowledge based on the blended teaching model taking the course Introduction to Big Data as the example. Hu Bibo [7] et al. constructed a new online and offline blended teaching model based on SPOC and peer-to-peer classroom, and took the Super Star Learning platform as the tool to achieve the balance between 'teacher-centered' and 'learning-centered'. They also took the 'Big Data processing and analysis of social website logs' in the course of 'Hadoop Big Data Technology' as the example to carry out the practice. Qin Haihuan [8] et al. have explored the blended teaching model of 'online independent learning, offline classroom learning, introduction of pair experiment and process assessment' in the course of Big Data Technology. Tang Guoming [9] et al. presented the flipped classroom design scheme of Introduction to Data Science based on the education cloud platform, and elaborated the specific implementation details of the course design scheme. Chu Fuyin [10] et al. explored the online and offline blended teaching mode of Big Data Development Technology from aspects of curriculum design, instructional design, evaluation system, and implementation effect.

Combined with the advantages of traditional teaching and network teaching methods, blended teaching not only gives full play to the leading role of teachers in the teaching process, but also uses the advantages of the information technology in the student-dominated 'online + offline' teaching [6]. On October 31, 2019, the Ministry of Education issued the Implementation Opinions on the Construction of First-class Undergraduate Courses, which clearly pointed out that the deep integration of modern information technology and teaching should be vigorously promoted, and flipped classroom and blended teaching should be carried out while supporting the organic combination of offline face-to-face teaching, so as to build blended 'golden courses' that integrates online courses with classroom teaching in universities. Considering the research results and practical experience of blended teaching of Big Data courses at home and abroad, combined with the reality of Chengdu University of Information Technology, this paper studies the blended teaching model of Big Data System and Technology, proposes the course content design, explores the course teaching method and establishes the course evaluation system under the blended teaching model.

2 Content Design of Big Data System and Technology Course Under the Blended Teaching Mode

The Big Data System and Technology course mainly introduces the basics of big data and the complete ecosystem of Hadoop to students, focusing on HDFS and MapReduce, explaining HDFS file management technology and the programming model based on MapReduce, and appropriately involving the big data ecosystem of HBase, Hive, Spark, etc. enable students to master the basic distributed file management techniques and the principles and implementation methods of high performance computing, and lay a good foundation for engineering work in big data-related professions.

The course requires students to understand the background and basic concepts of big data technology, master the deployment of Hadoop runtime environment, master the principle and shell operation of Hadoop distributed file system HDFS and access HDFS through JavaAPI, master the storage of big data files in HDFS, have the ability to modify the specific source code implementation of HDFS, the ability to develop MapReduce code, master MapReduce principle architecture, resource scheduling framework YARN operation principle, MapReduce application development methods, and analyze the specific process of MapReduce execution from the code point of view. The course also aims to make students have the ability to develop MapReduce code, have the ability to master Hadoop how to convert HDFS files into Key-Value for Map calls, the ability to master the internal operation and implementation details of MapReduce and transform MapReduce, the ability to build clusters and manage Hadoop, the ability to realize the programming implementation of sentiment analysis of big data sets by combining machine learning and using plain Bayesian classifier, the ability to understand the Hadoop framework ecosystem and its basic concepts and typical application scenarios, and learn to look at big data with innovative thinking, and learn the subsequent extension of big data.

Big Data System and Technology in Chengdu University of Information Engineering has 32 class hours, including 24 class hours of theory and 8 class hours of laboratory. It is impossible to cover a large amount of knowledge in a relatively short period of time

by traditional teaching methods alone, which also show the need for blended teaching. In accordance with the course objectives, teaching content under the blended teaching mode is reasonably designed based on the Internet quality course resources and the teaching resources recorded by ourselves.

1. The big data technology overview section mainly covers the era background, concept, impact, application, key technologies, computing models, industry of big data and the relationship between big data and cloud computing and the internet of things. Among them, the era of big data and the application of big data is arranged for online learning, mainly through the form of thematic video presentations.
2. The Hadoop part of big data processing architecture mainly covers Hadoop overview, Hadoop ecosystem and Hadoop installation and configuration. All content in this section is arranged for online learning, the Hadoop installation and configuration need to go through the teaching resources recorded by yourself, Detailed demonstration of how to install and configure Hadoop in pseudo-distributed and fully distributed environments.
3. Distributed storage HDFS part mainly includes Distributed File System, Introduction to HDFS, HDFS related concepts, HDFS architecture, HDFS storage principle, HDFS data reading and writing process, HDFS programming practice and HDFS cases, etc. Among them, Introduction to HDFS, HDFS programming practice and HDFS cases are arranged for online learning, HDFS Introduction to HDFS, Introduction to HDFS uses MOOC resources, and HDFS programming practices and HDFS cases use their own recorded teaching resources.
4. MapReduce part of the distributed computing framework mainly includes MapReduce overview, MapReduce workflow, MapReduce example analysis (WordCount), MapReduce specific applications and MapReduce programming practice. The MapReduce specific application and MapReduce programming practice are arranged for online learning, with MapReduce specific application using MOOC resources and the MapReduce programming practice using its own recorded teaching resources.
5. The rational part of big data mainly covers sentiment analysis, machine learning, plain Bayesian classifier and programming implementation of sentiment analysis. Among them, sentiment analysis and machine learning are arranged as online learning, utilizing their own recorded teaching resources.
6. The Hadoop ecosystem divide mainly includes HBase, Spark, Hive and other contents. Limited by the number of hours, this part is completed by course content, all arranged for online learning, with Internet resources as the main focus, supplemented by personal recorded resources.

3 Big Data System and Technology Blended Teaching Practice

Blended teaching is not equal to online teaching but is an integrated application of several teaching methods. The teaching of the Big Data Systems and Technology course uses a combination of flipped classroom, problem-oriented and case-driven methods, etc. And the teaching approach has changed significantly compared to the traditional model. The course in general divides the teaching into three different sessions before, during and

after the class [11], there are different tasks to be completed and different methods to be used in each session.

3.1 Pre-course Session

The pre-course plays a key role in the blended learning, which requires teachers to design, select and produce various resources for the course according to the course content, and choose a suitable platform. The course Big Data System and Technology uses the "Course Centre Online Teaching Platform' of the university (the teaching platform has functions such as course resource management, classroom playback, announcement management, assignment management, teaching interaction, student management, course setting and learning statistics). Prepared courseware, recorded videos, links to online resources, assignments, class exercises, test questions, etc. are organized according to a catalogue. Particular attention should be paid to the fact that teachers to record videos of course presentations so that students can understand the process, methods and format of the entire course implementation. Students are required to watch the videos, complete class exercises, assignments and discussions, and record any problems and gains in the learning process in relation to each teaching objective. The teaching platform can record students' learning time, exercise completion and homework completion, and teachers can keep track of students' knowledge acquisition. During the online learning process, students can communicate with teachers through the teaching interaction function of the platform, which facilitates the timely resolution of problems and avoids the accumulation of problems. In addition, students need to be grouped, during online learning, so that adequate discussions to prepare for the next step of the flipped classroom.

3.2 In-Class Sessions

The in-class is an essential stage in the blended learning model, which is concerned with the achievement of teaching objectives. In addition to explaining the designed offline content, during this phase, problems common to students' online learning need to be addressed in a focused manner. As the most common form of blended learning, the flipped classroom plays an extremely important role in enhancing the actual effectiveness of blended learning. Depending on the position of the online learning content in the overall course content system, the flipped classroom is interspersed with the offline lectures. Based on the grouping in the pre-course phase, the teacher randomly selects a subgroup, which chooses a representative to explain the relevant content. At the end of the lecture, the other subgroups can ask questions, which are jointly answered by the group members. The teachers make a brief comment and identify gaps and fill them in after the students' lecture. In addition to the flipped classroom, the course uses a problem-based approach, where each lesson is preceded by a question, followed by an explanation of the relevant knowledge, followed by a discussion among the students according to their sub-groups, leading them to think and develop solutions step by step.

The case study method is an open-ended and interactive new teaching method, which leads students to discuss these particular scenarios through the description of a specific educational situation. In the case study method, the teacher takes on more of a designer and motivator role, while students participate in the teaching through group discussions,

giving full play to students' learning initiative and enthusiasm, enhancing their analytical and problem-solving skills, and adequately reflecting the 'learning by doing' aspect of engineering education. The course Big Data Systems and Technology chose 'Sentiment discrimination based on Naïve Bayes classifier' as a comprehensive case study, and decomposed it into 'Sentiment data collection', 'Sentiment data storage', 'Sentiment classifier training', 'Sentiment classification discrimination', 'Prediction result analysis' and so on, according to the knowledge points of the course. This series of sub-cases is used throughout the theoretical and practical sessions of the course. Through the implementation of each sub-case, a more complete big data project will be finished finally. Through the implementation of each sub-case, a complete big data project is finally completed. By using the sub-cases, the boring theoretical concepts are avoided and the course content is easier to understand, thus achieving an organic integration of theory and practice.

3.3 After-School Session

At the end of the lesson, teachers need to summarize the lesson in time and find out the shortcomings for improvement next time. At the same time, the teaching platform's interactive teaching function is fully utilized to answer students' questions and queries in a timely manner, and these questions will be recorded, in the form of test questions to test the students' mastery of knowledge, to further understand the learning situation. Students can check their knowledge and improve their knowledge after class (the teaching platform will provide teaching resources that are not required by the syllabus, such as data visualisation, data pre-processing, NoSQL database, Flink, Pregel, Storm, etc.). In order to better develop students' practical hands-on skills and the ability to apply their knowledge in an integrated manner, teachers also combine big data competitions and actual projects to find complex cases for students and encourage those who are able to learn to try to complete them. These cases are highly complex, comprehensive and driven by practical needs, which are beneficial to cultivate talents to meet the needs of society.

4 Curriculum Assessment Reform

The evaluation of the course in hybrid teaching emphasizes the principles of combining process evaluation and result evaluation, the diversifying of evaluation subjects and evaluation indicators, and the comprehensive evaluating of evaluation content and evaluation methods. Therefore, the traditional way of usual grades + examination papers is no longer suitable, and the traditional assessment method needs to be reformed. The assessment needs to be fair and just, and can reasonably reflect the real situation of students. Combined with the curriculum to cultivate students' abilities in distributed storage programming and application, big data analysis and processing, and big data mining technology and application, the course assessment of Big Data System and Technology is transformed into an assessment of the learning process. The current course assessment grade mainly consists of the usual grades, experimental grades and final grades, each accounting for 50%. The usual grade includes attendance, learning of course resources,

completion of assignments, completion of class exercises, completion of tests, performance in the flipped classroom, exchange and discussion of problems, and completion of experiments. The final grade discards the traditional paper-based assessment and takes the form of a project defence, which mainly contains the completion of the project, the performance of the defence and the quality of the summary report. Table 1 shows the specific assessment standards of Big Data System and Technology course, including the assessment links, assessment content, score proportion and the assessment description.

Table 1. Assessment criteria for the Big Data System and Technology course [Self-drawn]

Examination session	Assessment content	Percentage of achievement	Appraisal instructions
Usual	Student Attendance	4%	Attendance is taken at least 4 times.
	Course Resources Learning	6%	Assessment is carried out in accordance with the learning progress data provided by the teaching platform.
	Assignment completion	2%	Based on the answers, the teaching platform is marked.
	Completion of classroom exercises	2%	Based on the answers, the teaching platform is marked.
	Test completion	3%	Based on the answers, the teaching platform is marked.
	Flipped classroom performance	7%	Evaluation is based on group presentations and answers to questions, according to group evaluation.
	Discussion on the exchange of questions	6%	Evaluation based on group classroom communication and discussion, according to group evaluation.
Experiment	Hadoop development environment configuration experiments	4%	Assessment is based on the completion of the experiments and the quality of the lab reports, with 50% each being awarded for the completion of the experiments and the quality of the lab reports.
	HDFS file access experiments	6%	Assessment is based on the completion of the experiments and the quality of the lab reports, with 50% each being awarded for the completion of the experiments and the quality of the lab reports.

(*continued*)

Table 1. (*continued*)

Examination session	Assessment content	Percentage of achievement	Appraisal instructions
	Experiments on sentiment discrimination based on Naïve Bayes classifier	10%	Assessment is based on the completion of the experiments and the quality of the lab reports, with 50% each being awarded for the completion of the experiments and the quality of the lab reports.
End of term	Project completion	30%	Working in groups, demonstrate the completed project and give a grade based on the completion of the project, and evaluate according to the group.
	Defending performance	10%	Students explain the work they have done individually and answer questions from the teacher, who gives a grade based on the actual situation. If the student is not involved in a specific realization, he or she is simply awarded a fail.
	Quality of the summary report	10%	The project summary report will be written in the prescribed format by the group, containing an abstract and body in English and Chinese, which needs to reflect the project analysis, project design, project implementation, problems and solutions, insights and experiences, etc. The grade will be given according to the quality of the report, and according to the group evaluation.

5 Conclusion

Big Data System and Technology, as a course of data science and big data and related majors, plays an important role in helping students master the basic knowledge of big data, apply the mainstream technical framework of big data, and cultivate students' thinking big data. Considering the characteristics of the course, such as complicated theoretical knowledge, high practical difficulty and limited teaching hours, combined with the actual situation of Chengdu University of Information Technology, the blended teaching mode is introduced into the teaching of Big Data System and Technology course. Firstly, through the course sorting, the online and offline contents were reasonably designed, and the teaching resources were clarified to support the teaching contents. Then the tasks to be completed in the three sessions before, during and after the class were determined, and how to apply the flipped classroom, problem-oriented method and case teaching method were explored. Finally, the assessment and evaluation system of Big Data System and Technology course focusing on the process assessment was discussed. The practice of blended teaching mode in the course has played an important role in

achieving the course objectives, improving teaching quality and cultivating students' abilities. The next step will be to continue to deepen the course teaching reform, optimize the course teaching content, and form course resources with its own characteristics.

Acknowledgement. This work was support by the Undergraduate Teaching Engineering Project of Chengdu University of Information Technology (JYJG2021048).

References

1. J. X. Wang, Z. W. Zhu, M. G. Li. The connotation, significance and implementing requirements of blended learning, Journal of Architectural Education in Institutions of Higher Learning, vol. 27, pp. 7-12, 2018.
2. K. K. He. Viewing the new development of educational technology theory from Blending Learning (I), E-education Research, pp. 1–6, 2004.
3. H. D. Li, H. Wu. Research on the mixed teaching quality evaluation system based on the whole process, China University Teaching, pp. 65–71, 91, 2021.
4. X. M. Zhu, J. Lei. Hybrid Teaching: Teaching Reform in the Era of "Internet plus", Digital Teaching in Primary and Secondary Schools, pp.4–7, 2018.
5. G. T. Luo, H. Yu. Research on the online and offline mixed teaching reform of Big Data Processing course, Education and Teaching Forum, pp. 121–124 , 2022.
6. L. L. Zhang, N. Tang. Content design and teaching method exploration of general education curriculum based on hybrid teaching mode -- Taking "Introduction to Big Data" as an example, Western China Quality Education, vol. 7, pp. 129-131, 2021.
7. B. B. Hu, M. Li, W. T. Zuo. Research on online and offline hybrid teaching mode based on SPOC -- Taking Hadoop Big Data Technology as an example, Computer Knowledge and Technology, vol. 18, pp. 139-140, 2022.
8. H. H. Qin, L. H. Lu, P. Lv, et al. Exploration and practice of mixed teaching of Big Data Technology course, Journal of Fujian Computer, vol. 38, pp. 110-113, 2022.
9. G. M. Tang, J. Lei, X. Zhao, et al. Flipped classroom design of Introduction to Data Science based on education cloud platform, Computer Education, pp. 18–22, 2019.
10. Y. F. Chu, Q. L. Zhang, J. Liu. Exploration and practice of online and offline hybrid teaching in Big Data Development Technology course, Computer Knowledge and Technology, vol. 16, pp. 110-111, 2020.
11. Y. Yang, J. Zhang. Exploration and practice of hybrid teaching of web front-end application development technology course, Journal of Higher Education, vol. 8, pp. 121-124, 2022.