Exploration and Practice of Big Data Literacy Cultivation Path for University Students*

Zhong Shifen Liu Rongqi Chen Honghong He Zhongxiu

School of Computer and Software Engineering
Xihua University，Chengdu，China
zshifen@qq.com, 394431956@qq.com

Abstract—In the big data era, a large number of students with big data thinking are needed. Keeping pace with the development of the times, the big data course for a small number of non-computer majors was offered in our university in the early stage, and then a teaching scheme of the big data general course for all students in our university was proposed. The teaching scheme includes thematic teaching content, blended teaching method, course assessment, and evaluation mechanism which focus on teaching process and ability training. The scheme has been practiced for 5 semesters. According to the survey data, 99.2% of the students have a general or even deep understanding of big data after learning the course. The survey data shows that the big data general course is very effective in cultivating students' big data literacy. It also can effectively improve students' big data application ability.

Keywords—*big data, data literacy, general course, blended teaching*

I. INTRODUCTION

Changes in the big data era are reflected in thinking, business and management. Big data thinking is conducive to the cultivation of interdisciplinary vision and lifelong learning ability. With the rapid development of information technology, data is growing exponentially. Big data thinking which integrates with industries, has a profound impact on the development of various disciplines, resulting in new fields of training and research, such as intelligent finance, intelligent manufacturing and intelligent writing and so on, as shown in Fig.1. Society has become increasingly reliant on data, making it necessary to ensure that all citizens are equipped with the skills needed to be data literate[1]. Big data has a wide range of applications. There is a surge in demand for not only big data technology professionals, but also compound talented people who understand big data knowledge. The big data policy is gradually extended from the comprehensive and overall planning to all major industries and sub sectors, and all kinds of big data employees is urgently needed. University students are the backbone of future scientific-technological development and social reform. It is imperative to cultivate university students' ability of data thinking and data analysis. How to cultivate these compound talented people for the society has become a major topic.

Yang Weiming *et al.* [2] propose dual-professional compound curriculum system of "big data + pedagogy" and "mathematics + big data" under the school-enterprise

collaborative education mechanism. Xiao Dawei *et al.* [3] give a training path under the background of new engineering.
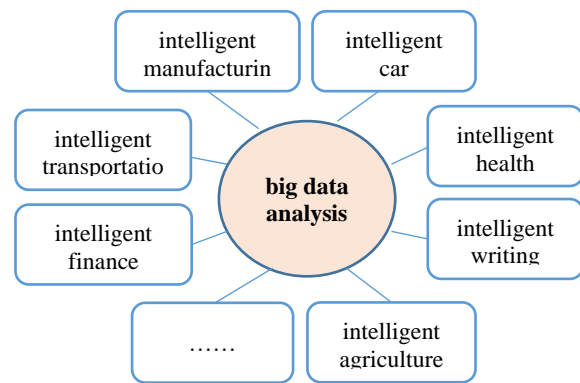

Fig. 1. Big data analysis training mode

Li Dalin et al. [4] propose that the ultimate goal of big data talents training is to cultivate excellent application-oriented talents, and give a diversified training scheme integrating school and enterprise resources. Gao Yanfang *et al.* [5] analyze and give the ways to cultivate the big data literacy of graduate students of management science and engineering in three aspects: teaching, scientific research and practice. Mark Frydenberg[6] explores approaches for integrating Big Data into the Information Systems curriculum. The training schemes or approaches proposed above are limited to big data professionals or interdisciplinary talents in some disciplines, and it is not universal to cultivate big data compuond talents for all majors. In 2016, Wang Wei [7] put forward the view that big data is suitable as a practical course of general education.

General education courses have a large audience and are generally offered in the form of large classes. Zhong Hao [8] mentions in the article that with the help of cloud teaching platform, we can solve the problems of large number of students in general education classes, the difficulty of process evaluation, less class hours, limited dissemination of general education concepts, loose organization, dull classroom, single assessment, low efficiency and so on. The research conclusion of Wu Yukun *et al.* [9] show that the flipped classroom teaching mode integrating big data analysis and mobile learning is better than the traditional teaching mode in classroom effect and test passing rate. Zhang CE *et al.* [10] conclude that online and offline blended teaching will become a trend through the multi-dimensional comparison between online teaching and classroom teaching.

Since 2016, our university has gradually carried out the teaching practice, and a set of big data general education training program was formed. The big data course for a small number of non-computer majors was offered in our university in the early stage, and then the big data general course for all students in our university was offered.

II. TRAINING OBJECTS AND OBJECTIVES OF BIG DATA GENERAL COURSE

A. Training Objects of Big Data General Course

In the article, Yang Runqian et al. [11] divide big data talents into three categories: technical, consulting and compound talented people. Among them, big data compound talented people refer to those who have big data thinking and professional knowledge of the work they are engaged in, who can give ideas and plans to support their own work by using data operation from the perspective of managers or their work, or assist big data technical talents to design big data application systems that meet their own work. All non-computer students can grow into big data compound talented people.

In a broad sense, big data compound talented people can be extended to comprehensive talents with big data literacy. Such people are needed in all walks of life and the demand is huge. How to cultivate them is a big challenge. In addition to being an introductory course for computer majors students, the main training object of the big data general course is all non-computer major students. It is of great significance to set up a gerneral course of big data.

B. Teaching Objectives of Big Data General Course

The teaching goal of the big data general course is to cultivate comprehensive talents with big data literacy. As an extension of information literacy, data literacy is also a necessary literacy for people in the era of big data, including data thinking, data ability, data security and ethics[12]. G. Jonathan [13] and others believe that data literacy education can not only cultivate people's sensitivity to data science, but also to data sociology, data politics and participation in public data infrastructure.

Big data literacy is an important part of scientific and humanistic literacy. In order to solve the limitations of data literacy training for students who were more specialized, the focus of big data general courses for undergraduates of all majors was no longer solely on cultivating students' big data skills and knowledge, but also on cultivating students' national feelings, data thinking, data ability, data security and ethical awareness. In terms of thinking enlightenment, it was important to guide students into the big data era, actively participate in the wave of big data reform, and enable students to have interdisciplinary thinking ability. In terms of data ability, it focused on cultivating students to understand the concept of big data, be familiar with big data applications, explored the combination point of big data and their own professional applications, and stimulated students' enthusiasm for innovation and entrepreneurship based on big data. In terms of big data security and ethical awareness, it was important to cultivate students' awareness of data security, keep calm thinking in the big data fever, correctly understand the security and ethical problems brought by big data technology, including privacy disclosure, information security, data gap, etc., and strove to make their behavior meet the requirements of big data ethics.

III. SPECIFIC IMPLEMENTATION OF THE COURSE

A. Build a Big Data General Education Talent Training Model

The cultivation of general talents in big data focused on thinking construction and application ability training, and the goal was to train innovative talents in the big data era who can solve their own professional and multi-disciplinary problems with big data thinking. As shown in Fig. 2, according to the teaching objectives of big data general education course, the training mode was divided into three levels of integration and coordination: thinking consciousness, knowledge and ability. Thinking consciousness focused on cultivating students' scientific spirit, big data thinking, big data security and big data ethics; Basic knowledge included basic concepts of big data, big data collection and storage, big data analysis and visualization, big data security and ethics, and big data applications, etc. In terms of application and innovation ability, it included the cultivation of literature review ability, project of big data and professional integration, big data application innovation, etc.

B. Construction of Hierarchical and Classified Teaching System

- Construction of hierarchical and classified teaching system

Our university is a comprehensive university integrating science, engineering, literature, economics and management, which provides theoretical and application support for each other. The big data general course focused on the organic combination of big data and various majors, taking into account the classification of the implementation of curriculum objectives and individualized teaching. In view of the fact that big data knowledge included not only universal contents such as data thinking and data application, but also professional contents such as data acquisition and data processing, the big data general education course was divided into two levels: basic and advanced courses. For liberal arts majors, students only needed to understand the concept of big data, the impact of big data on daily work, life and thinking, as well as the application of big data in different fields (especially their own fields), big data security and ethics. For majors with high data processing requirements such as information management and e-commerce, students could further expand their technical knowledge such as big data processing in addition to mastering the basic contents of big data general course.

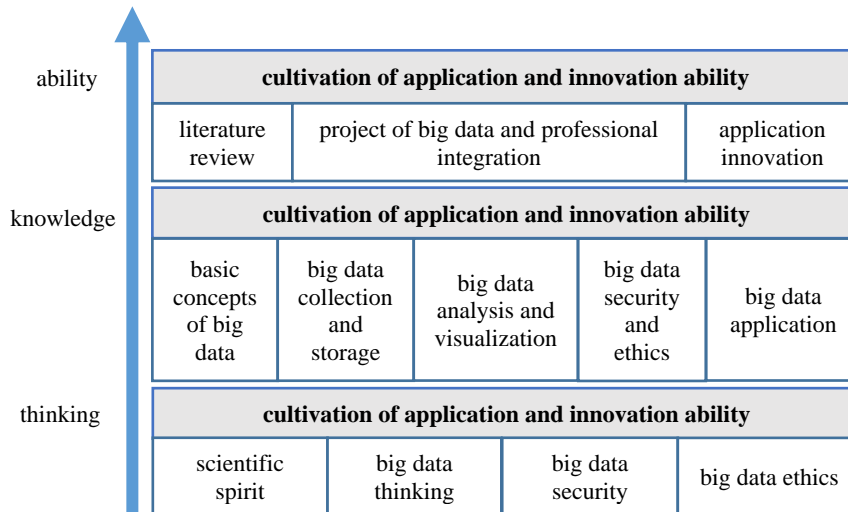| ability	cultivation of application and innovation ability		
	literature review	project of big data and professional integration	application innovation
knowledge	cultivation of application and innovation ability				
	basic concepts of big data	big data collection and storage	big data analysis and visualization	big data security and ethics	big data application
thinking	cultivation of application and innovation ability		
	scientific spirit	big data thinking	big data security	big data ethics

Fig. 2.   Three level integration collaborative training mode

- Design special teaching content

The popularization of big data is the future of big data. Only when people in all walks of life can use big data, the value of big data can be continuously enlarged and the industrial efficiency can be greatly improved. To enable all students to have big data thinking and big data literacy, to understand big data and big data application as well as basic computer knowledge, will better promote the development of all walks of life in the era of big data. In view of the diversity of the audience of students in the big data general education course, we designed the teaching content into multiple topics, including: meeting the big data wave, flashing data ocean, big data collection, big data storage, big data analysis, big data visualization, big data governance, big data security, big data ethics, big data application, etc. the specific contents and cases were appropriately modified in combination with the current situation in each lecture.

C. Using a Blended Online and Offline Teaching Method

Blended learning combined the advantages of traditional learning methods with the advantages of digital learning. It not only gave play to the leading role of teachers in guiding, enlightening and monitoring the teaching process, but also fully reflected the initiative, enthusiasm and creativity of students as the main body of the learning process. The big data general education course is a universal course with a large number of students, and it is a large class teaching form. The blended teaching method is more conducive to improve the teaching efficiency and quality.

- Blended teaching mode

The blended teaching mode of the big data general course included: students arranged online MOOC+SPOC learning independently; teachers flexibly arranged online live teaching, discussion and answering question according to teaching needs; teachers effectively organized offline classroom teaching, including divided-class, Case-based teaching, discussion-based teaching and other teaching methods; online discussion between students, teachers and students; offline and online assessment and evaluation, etc. For example, combining QQ group and cloud teaching platform to complete teaching work such as notification, announcement, video learning, discussion, and Q &A, and combining live broadcast platforms (such as Tencent conferences, QQ group classrooms, etc.) to conduct live teaching or real-time discussion and answering questions.

- Students' learning

Students' learning was divided into online diversified learning and offline classroom and extracurricular learning. When carrying out the big data general course, teachers uploaded learning materials and learning tasks to the cloud teaching platform in advance to guide students to preview relevant content in advance. Teaching materials included teaching schedule, teaching courseware, teaching video or MOOCs, cloud classroom related video, graphics, etc; Learning tasks could be weekly learning requirements and learning goals, as well as pre-arranged homework to be completed. By comparing the learning tasks and taking the completion of pre-class homework, in-class discussion and after-class test as the means to test the learning effect, students could achieve targeted previews and after-class summaries in advance. Through online learning, students entered the classroom with a good knowledge base, and arranged time independently after class to consolidate their knowledge, so that the traditional teaching could be better extended in time and space.

- Teacher teaching

Teacher teaching was divided into two types: offline and online. In order to fully mobilize the enthusiasm of students and avoid filling the classroom, the offline classroom teaching should adopt the form of split classroom as far as possible. Teachers only talked about the key points and difficult knowledge, and guided students to complete the learning of main knowledge by means of speech or discussion. Different from the conventional teaching, due to the assistance of wired

learning, in addition to explaining the key and difficult knowledge in the classroom, teachers could also focus on the online test questions and problems in homework, organized students to state their knowledge points, and guided students to discuss the problems encountered in online and offline learning.

- Teacher-student interaction

Cooperative and interactive learning methods is very effective in constructing knowledge and cultivating students' innovative ability. In order to promote students' effective learning, teachers carefully designed cooperative and interactive teaching links. With the help of QQ group, Tencent conference and intelligent cloud teaching platform, etc., teachers' teaching activity design could be more diverse, more interesting and easier to mobilize students' learning enthusiasm. In classroom teaching, teachers selectively use the convenience provided by the cloud teaching platform according to the teaching content, and activities such as voting questionnaires, on-site rush answers, brainstorming, discussions, etc., so as to activate the classroom atmosphere and make the boring classroom learning complete in a relaxed and pleasant atmosphere as far as possible.

IV. LEARNING FEEDBACK AND ASSESSMENT

A. Academic Feedback

Whether online or offline teaching, effective feedback of learning situation is an important link to improve teaching quality. The large-class teaching characteristics of big data general education course need to use various intelligent cloud teaching platforms to help collect learning information and dynamically adjust teaching strategies. The investigation of learning situation before class can help teachers understand students' learning fundation and learning conditions, so that teachers can adjust teaching methods and teaching plans appropriately according to their learning situation. The feedback of learning situation during teaching is helpful for teachers to master students' learning emotion and knowledge, and to appropriately adjust teaching priorities and teaching methods. The investigation and collection of opinions after the course will help teachers summarize teaching and improve it in the future. Using the functions such as "voting questionnaire" or "brainstorming" of the cloud teaching platform can implement different ways of learning situation surveys, and help teachers adaptively improve teaching progress and teaching methods. The intelligent cloud teaching platform can also automatically push some basic learning information, such as attendance, completion of activities, etc. Since all course development data are stored on the server, teachers and students can view the data at any time. Through these data, teachers can adaptively restructure teaching contents, adjust interactive links, and communicate with students. Students can also understand their learning in the class and realize self-supervision and motivation. Blended teaching with the help of multiple platforms can focus on the process more effectively than purely lecture-based teaching. By implementing blended teaching, students' learning efficiency can be significantly improved.

B. Assessment and Evaluation

With the development of society, what university students need to learn and how they learn are constantly changing, and teachers' evaluations and reports on students also need to change accordingly. The study of big data general courses not only includes the mastery of big data knowledge, but also covers "soft skills", "transversal skills" and "general skills" applicable to any field or category. Students need to develop lifelong learning skills. Making full use of learning process data can better evaluate and report on learning courses, especially the improvement of complex ability and general ability. Using the digital trace of learners' activities for learning evaluation can calculate scores in real time and greatly improve the timeliness of calculation and feedback in the course progress.

In order to better improve students' "soft skills", "horizontal skills" and "general ability", the traditional assessment model of "one paper at the end of the semester" has been reformed. An open and whole-process assessment method centered on ability training was implemented. It solved the problem that the traditional curriculum evaluation mechanism emphasized the result and ignores the process. The assessment and scoring methods of this course have also been reformed. The "five-five" ratio was implemented, that was, the total score of the course was composed of 50% of the usual performance score and 50% of the final course report score. Classroom or online discussion interaction, classroom questions and answers, questionnaires, chapter tests, video learning, and material reading were important evaluation indicators of course performance. The survey data showed that the big data general course was very effective in cultivating students' big data literacy.

C. Practical results

The evaluation dimension of data literacy included the sensitivity to data, the ability to collect data, the ability to analyze and process data, the ability to make decisions using data, and the critical thinking about data. Take the questionnaire survey conducted on some students who have chosen this course as an example. Among the 1,979 valid questionnaires about the knowledge of big data, 0.8% of the students were less sensitive to data after the course, 30.6% of them had basic knowledge, 53.1% of them had mastered it well, and 15.5% of them had in-depth understanding. The results of the valid questionnaires on data literacy are shown in Fig. 3. These questions included:

question 1：Able to keenly discover the existence of data information in life.

question 2：Can recognize and distinguish data format, type and other characteristics.

question 3：Understand the data itself.

question 4：Be aware of the significance of data in the corresponding field.

question 5：Be able to judge the accuracy of collected data and eliminate invalid data or obviously wrong data.

question 6：Be able to describe the statistic of data (e.g. mean, variance, etc.).

question 7：Can use visualization tools to visualize the data obtained.

question 8：The obtained data can be properly saved and backed up.

question 9：Be able to comply with the morals and ethics involved in data collection, utilization and sharing.

question 10：Can effectively protect data security, and protect data by means of laws when data is illegally infringed.
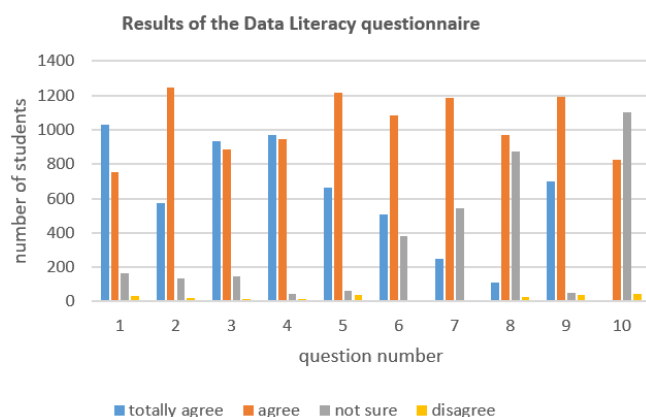

Fig. 3.   Results of the Data Literacy questionnaire

The survey data showed that the big data general course was very effective in cultivating students' big data literacy.Through our teaching mode of the big data general course, most students has understood what data is and the importance of data, had certain abilities of data collection, analysis, processing and decision-making, and had good sensitivity and professional cognition to the data in their professional fields.

V.   CONCLUDING REMARKS

In the big data era, society needs a large number of compound talented people with big data literacy. The main purpose of the big data general course is to cultivate all kinds of innovative talents who understand the concept of big data and have big data thinking. The big data general education course had a large number of learning objects with different professional backgrounds. It was an effective teaching implementation process to organize the teaching content according to the theme, stratify and classify training, and adopt a blended teaching method. The multi-platform-based blended

teaching method covered many teaching methods such as lecture-style, heuristic-style, participatory-style, and discussion-style. It integrated the experience of MOOC, SPOC, micro-classroom, and flipped classroom, and combined the pertinence and directness of offline classroom with that of online classroom. The combination of interesting teaching has greatly promote the overall improvement of students' cognition and skills. Practice has proved that the big data general course adopts a blended teaching method, combined with an assessment and evaluation mechanism that emphasized process and application ability, which was conducive to cultivating students' data thinking, data awareness and practical ability, so that students could more freely deal with the change of future world.

REFERENCES

[1]   A.Wolff, D.Gooch, J.J.C.Montaner, U.Rashid and G.kortuem. "Creating an understanding of data literacy for a data-driven society". The Journal of Community Informatics vol.12, 2016.

[2]   W.M.Yang, S.H.Wang, H.Wan and D.H.Yu. "Discussion on big data cross compound talented people training mode". University Education, vol. 6, pp. 31-33, 2020. （in Chinese）

[3]   D.W.Xiao, L.Q.Jiang and T.Li. "Research on the training path of applied big data talents under the background of new engineering". Computer Education. vol. 4, pp. 89-92, 2019. （in Chinese）

[4]   D.L.Li, W.Nu and Y.C.Liang. "Research on the training model of big data talents based on joint school-enterprise education". Software Guide, vol.19, pp. 164-167, 2020. （in Chinese）

[5]   Y.F.Gao and Y.X.Hu. "Research on the big data literacy cultivation for master degree dandidates of management science and engineering". Software Guide, vol. 19, pp.272-275, 2020. （in Chinese）

[6]   M.Frydenberg. "Introducing big data Concepts in an introductory technology Course". Information Systems Education Journal, vol.13. no.5. pp.12-23, 2015.

[7]   W.Wang. "Big data is suitable for general education practice". View Financial, vol. 000, pp.22-23, 2016. （in Chinese）

[8]   H.Zhong. "Research on cloud teaching of Japanese general elective course in application oriented universities -- taking Zhengzhou Business University as an example". University Education, vol.11, pp.127-129, 2019. （in Chinese）

[9]   Y.K.Wu and S.Z.Zhang. "Research on flipped classroom teaching mode integrating big data analysis and mobile learning". Journal of Educational Technology, vol.18, pp.40-42, 2019. （in Chinese）

[10]   C.Zhang, X.F.Xu, D.H.Chu, S.L.Gu and Z.Wang. "Reconstruction, integration and innovation -- analysis of blended teaching mode based on MOOC. Journal of Higher Education", vol.6, pp.1-8, 2021. （in Chinese）

[11]   R.Q.Yang and M.F.HAN. "Ability requirements and demand analysis of big data talents". Digital Technology &Application, vol.37, pp.2-7-208, 2019. （in Chinese）

[12]   R.H.Huang and B.Y.Li. "Data literacy education: expansion of information literacy instruction in the big data era". Document,Informaiton & Knowledge, vol.169, pp.23-31, 2016. （in Chinese）

[13]   J.Gray, C.Gerlitz, and L.Bounegru. "Data infrastructure literacy". Big Data & Society. vol.5, pp.1-13,2018.