Hierarchical and Diverse Cultivation of Data Thinking Capability in College Based on New High School Curriculum Standards

Jie Zhang, Ying Jin(✉), Li Zhang, Ye Tao, and Ping Yu

Department of Computer Science, Technology of Nanjing University, Nanjing, China
{zhangj,jinying}@nju.edu.cn

Abstract. Starting from the national demand for big data talent cultivation, universities have problems such as insufficient understanding and content mismatch in Cultivation of Data Thinking Capability. In particular, according to the cultivation requirements and contents of data thinking capability in the information technology curriculum of new high school Standards, this paper puts forward the reform of the cultivation system of data thinking capability in universities, including the exploration and Realization of course series, curriculum content, evaluation system and so on.

Keywords: new high school standards · college · data thinking capability

1 Background

In the "Action Plan for Promoting Big Data Development", the government proposed that big data is a "new driving force, new opportunity, and new method" [1] to "promote economic transformation", "reshape national competitive advantage", and "improve government governance capability". The government has also established a policy mechanism for personnel training." Big data talents are the talents engaged in big data related work, mainly including core talents engaged in research and development, analysis work, and composite talents with both industry background and big data skills." [2] The talent gap is expected to reach more than 2 million in 2025. Colleges and universities need to take the quality and effectiveness of talent development as the fundamental standard for working performance assessment, not forgetting the original intention of establishing moral education, remembering the mission of nurturing people for the Party and the

1. Association of Fundamental Computing Education in Chinese Universities, Teaching re-search project of 2021-College Computer Basic Curriculum Reform Based on new high school curriculum standards. 2. University Computer Course Teaching Steering Committee of the Ministry of Education, Research on the reform of university computer empowerment edu-cation in the new era project- Research on the reform of university computer empowerment education in New Liberal Arts. 3. Computer Education Research Association of Chinese Universities, Projects in the direction of education in 2021 - Research on the connection be-tween universities and K12 information technology teaching

W. Hong and Y. Weng (Eds.): ICCSE 2022, CCIS 1813, pp. 144–154, 2023.
https://doi.org/10.1007/978-981-99-2449-3_14

country, and not only focus on the in-depth training of professional talents, but also pay more attention to the exercise of composite talents.

The cultivation of computer technology talents is the demand of all industries in the information age, and the development of new technologies of cloud computing, big data, Internet of Things, and artificial intelligence has put forward the requirements for the cultivation of computer technology talents in-depth and breadth. However, the content of the university curriculum generally lags behind the development of the technology of the times, so the rapid development of computer information technology-related courses needs to keep pace with the times to reform and improve. The information technology curriculum standards for high school developed by the Ministry of Education were newly revised in 2017, and many courses were offered at the high school level to improve students' information technology knowledge and skills, enhance information awareness, develop computational thinking, improve digital learning and innovation skills, and establish a correct sense of worth and responsibilities in the information society. Students have already acquired information technology training in the high school period, but the learning objectives and knowledge composition in the high school do not support it to meet the requirements that society needs. Therefore, it is necessary to re-enhance students knowledge structure and skills through the cultivation of literacy, ability, and thinking at the university level.

Data is a resource, and the understanding and use of data affect all professions and industries. This paper discusses the development of data thinking capability from high school connection to college.

2 Cultivation of Data Thinking Capability in High School it Courses

The latest high school IT curriculum standards, i.e. the new standards, propose the core literacy of high school IT subjects, including four components: information awareness, computational thinking, digital learning and innovation, and information social responsibility. The four core literacies are influencing each other to support each other and jointly cultivate students' information literacy. The development of information awareness in the first one requires students to have sensitivity to information and judgment of its value. In turn, this aspect of competence needs to be acquired through proper analysis of data [3].

The high school information technology curriculum includes two compulsory courses, six optional compulsory courses, and two elective courses, totaling ten courses. There is one compulsory course named" Data and Computing", including data and information, data processing and application, algorithms and programming modules, and four optional compulsory courses named" Data and Data Structure"," Data Management and Analysis" and" Preliminary Artificial Intelligence". Data and Data Structures"," Data Management and Analysis", and" Preliminary Artificial Intelligence" are four courses related to data. Including data, data structure, data requirements analysis, data management, data analysis, and other contents.

The optional compulsory courses are more of a deep dive and expansion of the compulsory courses, starting not only with data modeling, data abstraction, and code implementation, but also learning to collect and use the right analysis and mining methods

from the massive amount of data, and understanding the powerful weapons for making scientific decisions. All need to recognize the importance of data, and the importance of data thinking capability development.

3 University Level Requirements for Data Thinking Capability

In the Basic Requirements for Teaching Basic University Computer Courses [4], it is mentioned that the development of basic university computer teaching should not only conform to the discipline's own rules of development, but also actively adapt to the needs of national economic and social development. China has made remarkable achievements in information technology construction and has established a large network environment, and is now in the historical development stage of strengthening independent innovation research to utilize the power of this large network. The deep integration of computer and information technology into various fields of economy and society, and the formation of a new form of economic development with the Internet as the infrastructure and realization environment are not only the macro strategy of the country, but also the basic vision that every university student should have when entering the society.

In such a historical period, basic university computer teaching faces great opportunity: it will take up the main task of cultivating the ability of computational thinking, which is one of the three pillars of scientific thinking; it will prepare the necessary knowledge and application ability for the cross-fertilization of computer disciplines with other disciplines, and it will cultivate information society citizens who have the literacy of computational thinking and are familiar with computer applications. Therefore, the development trend of basic university computer teaching in the next decade will show more significant characteristics.

The content of basic university computer education is divided into the following three fields: system platform and computing environment, algorithm foundation and program development, and data management and information processing. Among them, the field of data management and information processing: involves the basic techniques and methods of applying computer systems for data analysis and information processing, typically database technology, multimedia information processing technology, intelligent technology, etc. Its subfields are data organization and management, multimedia information processing, analysis, and decision making. The cources related with data are database technology, statistical analysis of data, data mining, artificial intelligence, etc. There is a certain continuity with the courses in high school.

Both the core literacy requirements of the new high school curriculum and the basic requirements of the university basic computer courses have the same goal in data thinking development requirements. But due to the knowledge structure and immature worldview of high school students, it is impossible to meet the goals of universities for talent development in terms of achieving the same literacy development. As shown in Table 1 below, 2 dimensions were selected to compare the requirements of high school standard IT and college basic computer courses.

In terms of curriculum content requirements, high schools are more likely to use" application cases" and" life cases" and then use words such as" experience"," feeling"," understanding" and" awareness" to describe the extent to which students have mastered

knowledge.”,” understanding”,” awareness” and other words to describe the degree of students’ mastery of knowledge. In contrast, the requirements of universities for similar contents are more focused on the mastery on” methods”,” strategies” and” theories”. Therefore, universities need to try to provide a more in-depth curriculum system for the cultivation of data thinking and computational thinking talents.

Table 1. I Comparison between the content requirements of high school it teaching and the requirements of university basic computer courses

content in High school standard	high school requirements	content in universities	universities requirements
Data and Computation Algorithm and Program Implementation	Outline the concepts and characteristics of algorithms from real-life examples, and represent simple algorithms using appropriate description methods and control structures. Acquire the basic knowledge of a programming language and implement simple algorithms using a programming language. Experience the basic process of programming through solving practical problems, feel the efficiency of algorithms, and master the methods of debugging and running programs. Describe the process of creation and development of programming languages and understand the functions and features of different types of programming languages	Algorithm Fundamentals and Program Development Algorithm and Programming	Understand the concept of algorithm and program; master the basic methods of iteration and recursion; understand the strategy of solving algorithms for typical problems; understand the concept and role of common data types and data structures; understand the idea of the modular design of programs, as well as the design ideas of module decomposition and compounding; understand the concepts of information encapsulation, interface, etc., as well as the ideas of reusability and reliability; understand the basic methods and ideas of program testing

(*continued*)

Table 1. (*continued*)

content in High school standard	high school requirements	content in universities	universities requirements
Data Management and Analysis Data Analysis	Understand common data analysis methods, such as comparative analysis, group analysis, average analysis and, correlation analysis, etc.; use appropriate data analysis tools in practice to analyze, present and, interpret data. Use digital learning to understand new developments in data management and analysis techniques; combine appropriate case studies to recognize the importance of data mining for problem solving and scientific decision-making in the information society	Data Management and Information Processing Analysis and Decision Making	Understanding of information representation and basic analysis methods in data analysis and decision making; statistical analysis, decision support systems, statistical analysis techniques, decision trees, association rules, an overview of artificial intelligence, an overview of data mining

4 Data Thinking Capability Cultivation System

Guided with the "3–3" undergraduate teaching reform, Nanjing University has been adjusting the basic computer education curriculum system after several rounds of reform, aiming to better align with the undergraduate teaching mode. Basic computer education to cultivate computational thinking as a whole extends to different fields of computing, in which the cultivation of data thinking has become one of the most important contents and goals of basic computer education in our university. This paper focuses on several aspects of the course series, course content, and evaluation system.

4.1 Course Series

For the needs of the students at different levels, we arrange general basic courses, elective courses, innovation and entrepreneurship courses in the data thinking course series. The basic course is offered as a compulsory course for science and engineering departments as well as arts and science departments, and students from art departments are encouraged to take it as an elective. Figure 1 shows the proportion of the students from the departments that arrange the Python course as elective course, in which includes all kinds of departments of arts and sciences in the university.

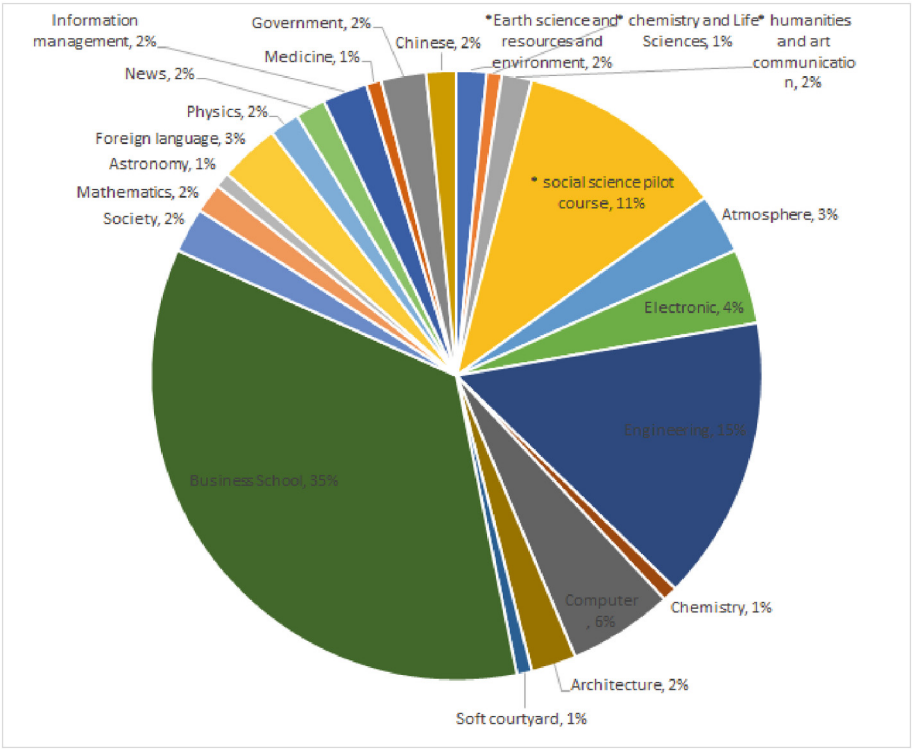

Fig. 1. The proportion of cross-platform course enrolment in Python programming courses

Elective courses are offered to all students, in principle, regardless of arts departments or sciences departments. But generally, students will choose courses that interest them or can improve their capability in depth and breadth. The elective courses related to data thinking development include Database Technology, Data Science, Data Processing Using Python, etc. Most of them are the extension for the content of the basic courses.

The innovation and entrepreneurship course is an important part of the "five-in-one" innovation and entrepreneurship teaching system of Nanjing University, and it is the main channel to promote innovation and entrepreneurship education in the university. It is offered to all students in Nanjing University. Figure 2 shows the proportion of students in the course, mainly in science and technology, including students from computer science and School of AI.

The curriculum series meets the requirements of talent cultivation in Nanjing University. The first stage is the general education stage for freshmen. The second and third years education focus on the specialization cultivation, and the fourth year for diversification cultivation.

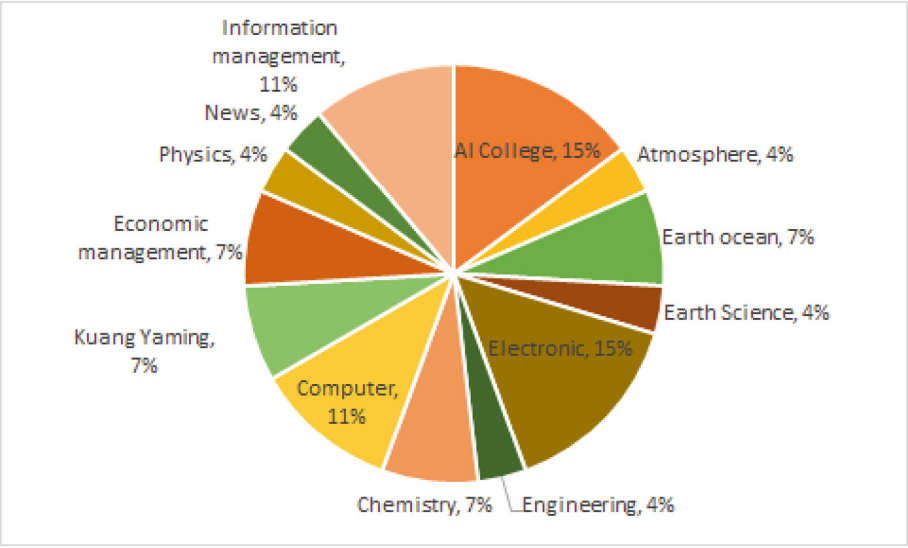

Fig. 2. The proportion of students in Introduction to Big Data and Python Implementation course

4.2 Content Level

The course content is based on the different courses in the curriculum system to develop a hierarchical setting, as shown in Table 2.

As a basic course, Python Programming takes the basic learning requirements of programming languages as a starting point. In addition, the using of scientific data processing libraries and the learning of basic methods of data processing are also added to this course. The aim is to enable the students to use Python to do data analysis and to support their research. In this course, we use IDLE or anaconda as the programing environment.

The elective courses on data processing in Python and R language focus on the process of data processing and analysis, i.e., data acquisition, data statistics, data mining. The course" Data Processing Using Python" focuses on the implementation of Python in the whole process of data processing, from data acquisition locally or from the web, how to represent data, to pre-processing, exploring, analyzing and visualizing data, and finally to representing and processing data through a simple GUI interface. In this course, we use engineering platform tools such as pycharm, vscode, etc.

In the innovation and entrepreneurship course Big Data Introduction and Python Implementation, which opens for the whole school, the course focuses more on industrial implementation and application. It includes the overview of big data, big data technology, big data analysis process, Python implementation, platform practice, big data industry application, and big data infrastructure. Both theoretical knowledge and cases are included in this course. It also includes practice on industry big data platform with real case study. Combining with industry hotspots and frontiers, this course enhances students' understanding and experience of innovation and entrepreneurship. Through the cultivation of the ability of comprehensive application of knowledge, students will be prepared to go deeper into the industry and solve practical problems. The curriculum resources are built in collaboration with external forces and based on industry-university

Table 2. Comparison of course content and tool platforms for the three levels

Course Level	Course Name	Course content related to data processing	Use of tools or platforms
Basic Course	Python Programming	SciPy, the scientific computing ecosystem; numpy library, pandas library, matplotlib library	IDLE, anaconda, etc
Elective Courses	Data Processing Using Python	Data acquisition: Local and Network Data Acquisition; Powerful Data Structures and Python Extension Libraries: extension libraries SciPy, numpy, pandas; Python data statistics and mining: Python convenient data acquisition and preprocessing, Python data statistics mining and applications (including cluster analysis, Python in science and technology, and humanities and social sciences applications). Object-oriented and graphical user interfaces (including the use of GUI libraries)	anaconda, pycharm, VS code, etc
Innovation and Entrepreneurship Class	Introduction to Big Data and Python Implementation	Big data overview, big data technology, big data analysis process, Python implementation, platform practice, big data industry applications, big data infrastructure	pycharm, VS code, FineBI, pandasBI, etc

research projects. In this course, besides of Python, students need to choose data analysis platforms and visualization platforms that can be practically applied in industrial applications, such as FineBI, etc.

For students who have spare capacity for learning, they are organized into teams to discuss the scientific research based on real data. The outcome should also be linked

with industry demands. Students are also encouraged to attend college student innovation competition programs. Then we could study from the competition program to refine the course designing.

The hierarchical diversification of course contents is set following the theory of talent cultivation in Nanjing University. In the diversified cultivation stage, students need to be divided into three directions - the specialization cultivation direction, the composite talent cultivation direction, and the innovative for employment direction. They are each designed for the students who are willing to continue to study along with their current major, students who want to study across the major, and students who want to start their own business in the future.

The whole series of courses and all the levels of course content for the cultivation of data thinking for non-CS major students have been set to improve the system of talent cultivation, which is in line with the training objectives of computer basic education for students to take basic courses. The courses also help the hierarchical students to carry out targeted improvement, which is in line with the objectives of cultivating inter-disciplinary talents, and enable students on innovation and entrepreneurship.

4.3 Evaluation System

The evaluation system is designed to support the hierarchy for different courses. The evaluation of the basic course examines the fundamentals, so it is divided into closed-book exams, project design, and course assignments. It examines not only the mastery of basic knowledge, but also the ability to solve practical problems with teamwork. The course project mainly focuses on examining students ability in data processing.

For example Fig. 3, students chose instant noodles as the research object and found over 3,000 instant noodles data from Japan, China, and Korea. The analyzed the data in multiple dimensions such as flavor, packaging, and origin distribution. Then they created word cloud based on the flavor level of different instant noodles types. Students analyze the data by their interests and from real life scenario, and finally visualize the results. Thus helps them to fully understand how to use Python to support real-life problems.

Fig. 3. Instant noodle types word cloud by favor level

The elective course evaluation is also contains closed-book exams and project design, but in which the requirements for project design have been increased. The students not only need to implement the data processing, but also to master some algorithms such as data mining or artificial intelligence machine learning.

Since the goal of the Innovation and Entrepreneurship class is to cultivate innovative and entrepreneurial talents, the evaluation of the course is mainly based on the practical assignments about Python programing and the using of big data platforms. It also requires the students to study the papers on the application of big data in related industries, and finish the course paper for evaluation. More attention is paid to students' understanding and practice of practical cases of industry applications.

For example, the students used the SVM method for customer churn prediction based on a cleaned dataset, which contains thousands of cell phone program customer churn information in Fig. 4. And then analyzed the impact of different parameters on overfitting, as shown in the Fig. 5 below.

```
# 3. Training SVM classifier
classifier = svm.SVC(C=2,kernel='rbf',gamma=0.01)
classifier.fit(train_data, train_label.ravel())
# ravel function defaults to row order first when downscaling. ravel function
does not affect the result without adding, but there will be a red warning.
# 4. Calculate the classification accuracy
print("training set: {:.2%}".format(classifier.score(train_data, train_label)))
print("Test set: {:.2%}".format(classifier.score(test_data, test_label)))
```

Fig. 4. Part of the code of the project in the student's thesis

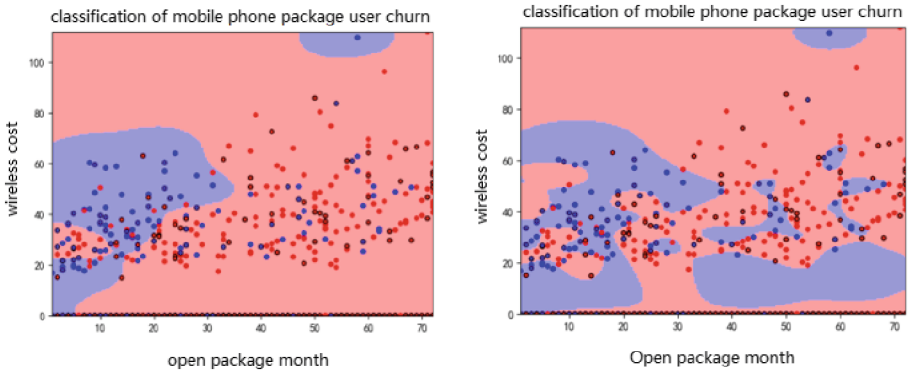

Fig. 5. Comparison of some visualization results of the projects in the student's thesis

The evaluation of university students' innovation and entrepreneurship projects is more open. Its evaluated by comparing the result with the goal set when applying the project. Thus to improve the real-life problem-solving capability through research-based learning approach.

5 Summary

The whole series of courses and all the levels of course content for the cultivation of data thinking for non-CS students have been set to improve the system of talent cultivation, which is in line with the training objectives of computer basic education for students to take basic courses, but also for the hierarchical students to carry out targeted improvement, which is in line with the objectives of cultivating interdisciplinary talents, and enable students on innovation and entrepreneurship.

However, there is still room for improvement in the overall course content design, especially how to deal with the gap between the content setting and the actual ability of students.

References

1. An action outline to promote big data development (GF [2015] No. 50)
2. China's big data industry development white paper (2019)
3. The 2017 edition of high School Information Technology Curriculum Standards
4. Basic Requirements for the teaching of College Computer Basic course 2015. Higher Education Press, Beijing (2015)
5. Fuliang, G., Gang, Z., Yongjie, L., Liangzhong, C., Hui, G.: Practical exploration on the curriculum reform of college computer foundation under the new syllabus. Comput. Eng. Sci. **41**(S01), 4 (2019)
6. Zhang, J., Jin, Y., Tao, Y.: Exploration on the construction of big data basic course for non computer major. In: 2020 15th International Conference on Computer Science and Education (ICCSE) 2020
7. Meijuan, W., Hui, L., Jingli, H., Changyou, Z., Yuanyuan, J.: Research and practice of college computer basic curriculum reform based on computational thinking. Comput. Educ. **3**, 5 (2020)
8. Yu, L.: Reform and practice of public computer basic curriculum based on innovative talent training mode. Educ. Modernization **7**(17), 57–59 (2020)
9. Yong, O., Hong, L.: Reform and practice of college computer basic curriculum under the background of engineering education certification. Comput. Educ. **4,** 4 (2019)
10. yunyun, W., Zenggang, X., Xia, K., Youfeng, L.: (2021)
11. Exploration on the reform of student - centered basic computer course Computer education (10), 5