

Research on the Course System of Data Science and Engineering Major

1st Zhenping Qiang

College of Big Data and Intelligent Engineering
Southwest Forestry University
Kunming, China
qzp@swfu.edu.cn

2nd Fei Dai

College of Big Data and Intelligent Engineering
Southwest Forestry University
Kunming, China
daifei@swfu.edu.cn

3rd Hong Lin

College of Big Data and Intelligent Engineering
Southwest Forestry University
Kunming, China
linh1226@swfu.edu.cn

4th Yueyu Dong

College of Big Data and Intelligent Engineering
Southwest Forestry University
Kunming, China
dongyueyu@swfu.edu.cn

Abstract—With the convergence and integration of information technology and human production and life, data has grown rapidly, and data has become the basic strategic resources of the country. It is followed by the construction of data science and engineering major in various universities, and this major is related to the disciplines of data science, computer science, mathematics and engineering management, and also a practice-oriented discipline. These reasons make very high demands on the training of talents in the data science and engineering major. Based on analysis features of the data science and engineering major and the goals that the professional needs, in this paper we discuss the basic skills that the talents should have, and pushes back to the construction of the professional curriculum system of data science and engineering major. Through the analysis method of this paper, we have completed to setting the course system of data science and engineering major and discussed the methods and the ways in the implementation of the curriculum system.

Keywords—course system, data science and engineering major, curriculum system

I. INTRODUCTION

Since 2012, data has been discussed as a “critical new form of economic currency”^[1]. In recent years, the convergence of information technology and economic society has led to rapid data growth, data has become a basic strategic resource for each country, and businesses in all industries are beginning to take advantage of the tremendous growth in data and new big data technologies to analyze and derive value from them. Demand for skilled data scientists continues to be sky-high, with IBM recently predicting that there will be a 28% increase in the number of employed data scientists in the next two years. In essence, with the advent of the era of big data and the continuous development of information technology, data science has increasingly become the most important technology in the modern era. These realistic requirements put forward new and specific requirements for the training of data science students. This article will focus on the content of data science teaching as a specific research content, combined with the experience of training students in recent years to give the planning of the data science direction curriculum system.

The data science major is mainly based on computer science and statistical analysis for the era of big data, with data science and big data technology as the research field, it

also requires the domain knowledge of economics, finance, data structure, data computing intelligence, information processing, programming, data analysis and other related disciplines. Through the training, the students are required to have the ability to promote and lead the future global Internet, cloud computing, artificial intelligence, and big data technology in various fields, while having strong practical ability, cross-cultural communication ability and interdisciplinary research, design, and development capabilities.^[2]

After clarifying the society's demand for talents, it is necessary to formulate student training programs(course system) in a targeted manner. The establishment of a curriculum system for data science and engineering major plays a vital role in cultivating students' knowledge and their successful job-hunting. In addition, the course system and course content are also the core of cultivating talent quality and improving the quality of teaching.^[3,4]

Specifically, the curriculum system refers to the system that combines the various components of the curriculum under the guidance of certain educational values, so that each curriculum element can be unified in the dynamic process to point to the realization of the objectives of the curriculum system. Therefore, the curriculum system planning of this paper mainly solves three problems: What is the training goal? What courses are set up? How are the courses integrated?

II. METHODOLOGY

A. Training goal

The most basic requirement for the cultivation of talents is to cultivate all-round development of moral, intellectual, physical and aesthetic, such as good psychology, physical quality, healthy body. Specific to data science and engineering talents also requires strong professional skills, business communication skills, innovation capabilities, etc., and for professional competence, that is, students are required to grasp the basic knowledge, theory and technology of data analysis, including basic knowledge of mathematics and computer science for big data applications, data analysis and processing, basic theory of business decision-making, basic methods and basic skills.

In order to understand the basic skills of students in data science and engineering major, let's start with understanding

the Data Science Lifecycle^[5]. As shown in Fig. 1, Data Science Lifecycle includes seven iterative steps, that is business understanding, data mining, data cleaning, data exploration, feature engineering, predictive modeling, and data visualization.

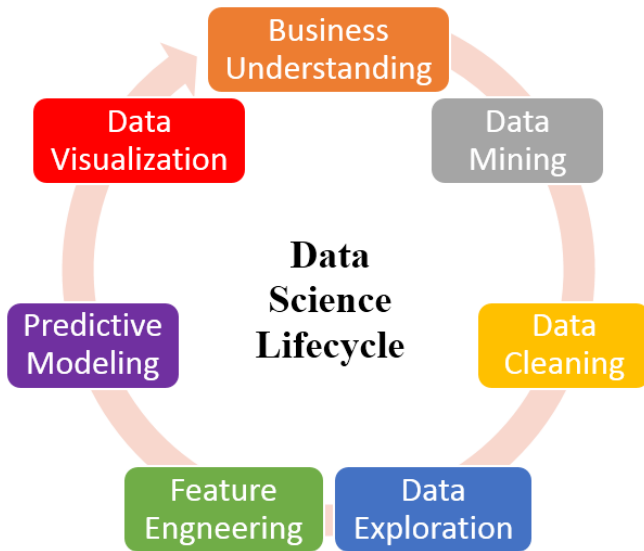


Fig. 1. Data Science Lifecycle, which includes seven iterative steps.

To complete data processing, in summary, graduates of data science and engineering major should have the following basic skills:

1) *Have mature data thinking ability*

The cultivation of data thinking ability requires students to learn the basic knowledge of computer science^[6], related mathematics and statistical knowledge, so that students can understand the contents of data, information, information representation, etc., and have a good insight into the data.

2) *Have skills in data processing*

Data processing skills are the key to the training of students' skills, which include the following aspects:

a) *Information mining ability*

The characteristics of big data are the large amount of data, the rapid generation of data, the diversity of data, etc. Although in recent years, deep learning and other technologies have reduced the requirements of data mining for original data, in the process of students' training, we must let Students learn how to remove unwanted data from the data and dig out the useful information they need to analyze.

b) *Information processing capability*

Information science is essentially the science of data processing, and the process of information processing includes the receiving, storage, transformation, transmission and distribution of information, which involves the application of digital signal systems, computer networks, and tools and technologies used in data processing.

c) *Computer coding ability*

After understanding the basic concepts of information and understanding the basic processing of information, the most critical teaching content of the data science and engineering major is the use of computers to complete the processing of data, this is the ability of computer coding. The input data of information processing unit may be text, language, image or more complex data information. And

after being processed by computer programs or algorithms, the value in the data can be exerted, therefore, this requires the graduated students to have computer programming skills.

3) *Have ability to work in teams*

The process of data processing, especially in big data environment, will inevitably require a team to complete. This puts forward the requirement of cultivating teamwork ability for the training of students. Specifically, in the teaching, students can be taught the knowledge of software project management, code sharing management and so on, and a specific task is accomplished by means of a teamwork collaboration to develop students' abilities in this area. In paper [7], the author describes how to implement Big Data science education with a case study of a project-focused introductory course.

4) *Some interdisciplinary knowledge is needed*

One of the biggest features of data science is cross-industry. As big data penetrates into various industries, big data practitioners often have multiple roles and need to master both data technology and business knowledge. For example, a data analyst must have the skills of data analysis, data mining and machine learning, as well as knowledge and skills in marketing, business models, and data products. An analyst who can do business data analysis and machine learning and engineering development has become a data scientist, and this is exactly the talent that is most needed now. The cultivation of interdisciplinary knowledge is difficult to achieve only through the teaching of data science and engineering major's courses. A good strategy is to consider combining the backgrounds of different schools to carry out teaching in this area. For example, for a college with a background in forestry, some courses in forestry-related basic knowledge and application systems are a good choice.

From the above analysis, we can get the relationship between data science & engineering major and other disciplines as shown in Fig. 2.

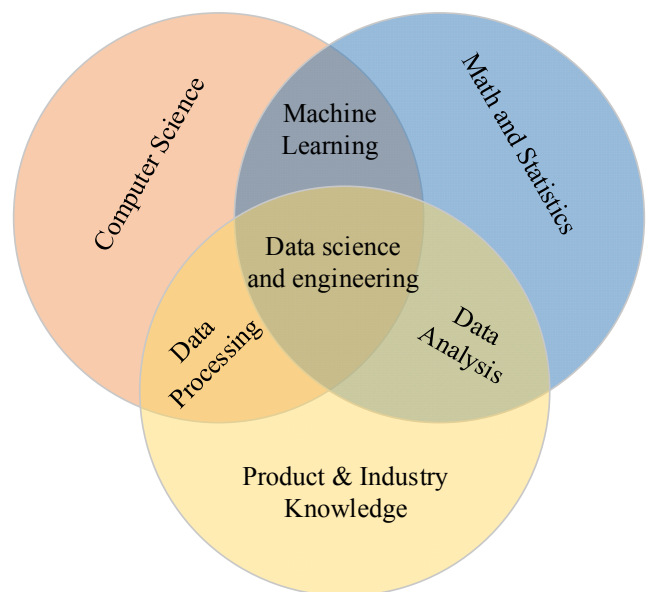


Fig. 2. Relationship between data science & engineering major and other disciplines

As Josh Wills gave a talk at Airbnb about the life of a data scientist, "A data scientist is someone who is better at

statistics than any software engineer and better at software engineering than any statistician.”^[8] It can be seen that the foundation of data science is the data processing methods involved in computer science, the foundation of the natural sciences—mathematics and statistics, and the ability to understand the data and problems in the field of substantive expertise. Accordingly, the professional course of Data Science and Engineering Major can be divided into two parts: basic theoretical level courses and the skill level courses. The basic theory courses are the foundation, which mainly includes data science, computer science and technology, mathematics and statistics and project management. In study of the conduction to determine factors that promote success in an introductory college computer science course. Bernard Marr^[9] also illustrates the importance of the mathematics in computer science major. The skill courses include data collection, data storage, data mining, data cleaning, data visualization, data modeling, data platform construction and application. The corresponding relationship between them is shown in Fig. 3.

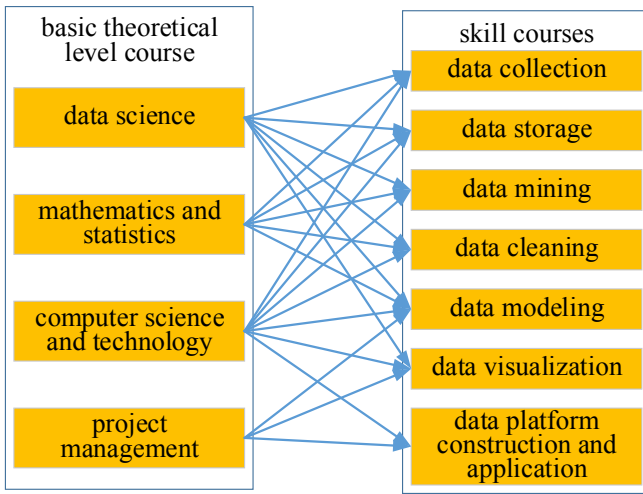


Fig. 3. The corresponding relationship between the basic theory courses and skill level courses of Data Science and Engineering Major

All basic theoretical level courses and skill level courses will be completed in four academic years. The relevant curriculum settings and academic year distribution are shown in Table I.

TABLE I. COURSE SETTING AND DISTRIBUTION OF DATA SCIENCE AND ENGINEERING MAJOR

Academic year	Courses
The first academic year	Introduction to Data Science , Calculus, Linear Algebra , Programming Foundations , Python programming
The second academic year	Probability Theory , Discrete Math, Intro to Statistics , Data Structures/Algorithms , Computer network, Big data foundation, Artificial Intelligence
The third academic year	Computer Systems and Architecture, Databases , Statistical Modeling/Regression , Machine Learning/Data Mining , Theory of Statistics
The fourth academic year	Data collection and web crawlers, Data Visualization Technology, Big data development platform, Capstone Course with Data Experience and Projects, Ethics

These courses cover the bare necessities of the material required for a Data Science and Engineering major. Among them, 10 courses highlighted in bold form are more

important. In addition, the Capstone Course with Data Experience and Projects in fourth academic year, require students to be able to solve practical problems in a specific field of substantive expertise, The settings can refer to the main professional background of the university where the major is built.

III. DISCUSSION

Data Science and Engineering major belongs to a major with obvious engineering characteristics. It needs computer technology, statistical theory, and management science to solve problems in specific industries. This puts specific requirements on the teaching process of the course:

1) *We need to pay more attention to the cultivation of students' thinking ability*

In the curriculum system setting, we integrate data science and computer ability. And take data science thinking and computer operation ability as the foundation. Further, we focus on the design of teaching content from the aspects of data processing, data value utilization and problem solving methods. For example, a freshmen can use the Python program examples to understand the knowledge of image processing and to understand the nature of data processing.

2) *We focus on cultivating students' practical ability*

The cultivation of practical ability is mainly embodied in the design of experimental content in the course teaching. First of all, in class period allocated, it is necessary to increase the proportion of practical teaching or experimental teaching. Secondly, it is suggested to adopt a heuristic teaching method to inspire students' self-learning ability.

3) *Learning mode reform must be carried out*

A major feature of data science is that its related technologies and methods are developing very fast, which requires practitioners in this field must have self-learning capabilities. In paper [10], Chamillard, A. T. et al. proved that the cultivation of learning style is important for students to understand new content. This requires the teacher to train the students' self-learning ability during the teaching process. We can start from the promotion of learning interests, through the form of task groups combined with task objectives to exercise. For example, in the data mining course, clustering, classification, association, statistical analysis, etc. can be used to guide students to develop self-learning ability according to algorithm decomposition tasks.

IV. CONCLUSION

In general, data science and engineering major, as a highly practical discipline, must pay attention to the study of students' practical ability in subject teaching, pay equal attention to theoretical teaching and practical teaching, and build a sound talent training plan. Construct a suitable teaching platform for data processing technology and application practice. When setting up the teaching system, we need to cover two aspects: basic theory and professional application. We can also fully examine the actual resources of the universities where the major is established, and expand the channels of practical teaching.

Through the research of this paper, we have developed a detailed professional curriculum system, which clarifies the relationship between different courses. It has a clear reference for the cultivation of data science and engineering major.

ACKNOWLEDGMENT

This work was supported by the Key Scientific Research Foundation Project of Southwest Forestry University (Grant No.111827) and the Innovation Promotion Fund of Science and Technology Development Center of Ministry of Education(2018A01002).

REFERENCES

- [1] World Economic Forum Briefing (2012). Big Data, Big Impact: new possibilities for international development, January[EB/OL]. http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf
- [2] Chaturapruet S, Dee T, Johari R, et al. How a data-driven course planning tool affects college students' GPA: evidence from two field experiments[C], Proceedings of the Fifth Annual ACM Conference on Learning at Scale, No. 63. London, United Kingdom, June 26-28, 2018.
- [3] Zhang Zheng, Zhou Tina, Hai Jinxing, et al. Being employment-oriented, strengthening engineering quality-talent training program and course system reform and practice of food science and engineering major [J]. Value Engineering, 2011, 30(30), pp. 174-175.
- [4] Cai-jun X U. Preliminary Study of Innovating Courses System of Geodesy and Geomatics [J]. Bulletin of Surveying and Mapping, 2007, 11.
- [5] Understanding the Data Science Lifecycle[EB/OL], <http://sudeep.co/data-science/Understanding-the-Data-Science-Lifecycle/>
- [6] Baumer, Ben. A data science course for undergraduates: Thinking with data[J]. The American Statistician, 2015, 69(4), pp. 334-342.
- [7] Saltz, Jeffrey, and Robert Heckman. Big Data science education: A case study of a project-focused introductory course[J]. Themes in science and technology education, 2016, 8(2), pp. 85-94.
- [8] what is a data scientist[EB/OL], <https://datascopeanalytics.com/blog/what-is-a-data-scientist/>
- [9] Bernard Marr, The 9 Best Free Online Big Data And Data Science Courses[EB/OL], <https://www.forbes.com/sites/bernardmarr/2017/06/06/the-9-best-free-online-big-data-and-data-science-courses/>
- [10] Eason G, Noble B, Sneddon I N. On certain integrals of Lipschitz-Hankel type involving products of Bessel functions[J]. Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, 1955, 247(935), pp. 529-551.