UIUC iSchool Data Science Course Cluster Research Post-Print

Authors: Yang Ruixian, Wan Jiaqi

Date: 2023-04-01T16:15:59+00:00

Abstract

[Purpose/Significance] This study investigates the current status of data science curriculum cluster construction, focusing on data science talent cultivation programs, to provide reference and guidance for data science teaching practices in information schools of Chinese universities. [Method/Process] Based on the data science curriculum practices of the School of Information Sciences at UIUC (University of Illinois at Urbana-Champaign), we first investigated the names, descriptions, credit hours, teaching formats, instructors, and target audiences of data science-related courses in this school. We then systematically classified and comparatively analyzed the curriculum cluster from four aspects: target student type, teaching format, degree of teaching collaboration, and course content. Finally, we proposed several recommendations for data science curriculum construction in Chinese universities. [Results/Conclusions] The UIUC data science curriculum cluster can be divided into six major categories, targeting students at undergraduate, master's, and doctoral stages, adopting a blended teaching mode combining online and offline methods, conducting teaching through faculty collaboration, with teaching content that closely follows the market demands of data science positions. Therefore, Chinese universities should strengthen cultivation continuity, enrich teaching innovation, enhance faculty teaching collaboration, and enhance research direction completeness in the data science field.

Full Text

Preamble

Investigation and Research on the Construction of Data Science Course Groups at UIUC iSchool

Yang Ruixian, Wan Jiaqi School of Information Management, Zhengzhou University, Zhengzhou 450001

Abstract: [Purpose/Significance] This study investigates the current state of data science curriculum group construction, focusing on data science talent training programs to provide references for data science teaching practices in Chinese university information schools. [Method/Process] Based on the data science curriculum practices at the University of Illinois at Urbana-Champaign (UIUC) School of Information Sciences (iSchool), we first examined the names, descriptions, credit hours, teaching formats, instructors, and target audiences of data science-related courses. We then systematically classified and comparatively analyzed the curriculum groups from four perspectives: target student types, teaching formats, degree of instructional collaboration, and course content. Finally, we propose several recommendations for data science curriculum construction in Chinese universities. [Result/Conclusion] The UIUC data science curriculum can be divided into six categories, catering to undergraduate, master's, and doctoral students through blended teaching methods combining online and offline instruction, with collaborative teaching by faculty and content that closely follows data science job market demands. Therefore, Chinese universities should strengthen educational continuity, enrich teaching innovation, enhance faculty collaboration, and improve research direction completeness in the data science field.

Keywords: data science; UIUC; iSchool; curriculum construction; curriculum reform

The Fourth Industrial Revolution has promoted the transformation of social development models, resulting in a shortage of talent in big data basic research, product development, and business applications [1]. According to a professional data talent education industry ecosystem report released by TalkingData, China will face a shortage of 2 million data science talents by 2025 [2]. Current domestic and international data science job skill requirements exhibit characteristics of "specialization with broad knowledge," with high demand and intense competition for high-end talent. The Ministry of Industry and Information Technology's "Big Data Industry Development Plan (2016-2020)" clearly states that the construction of big data industry talent teams urgently needs strengthening.

To adapt to the trends of the times and meet national construction needs, the Ministry of Education first added the undergraduate major "Data Science and Big Data Technology" in 2016, with 196 new major points added in 2019 alone and 25 new major points for "Big Data Management and Application" [3]. Therefore, exploring the current state of data science curriculum construction, optimizing data science talent training programs, and establishing a data science education system that meets social and market demands is imperative. The University of Illinois at Urbana-Champaign (UIUC) School of Information Sciences (iSchool) has long been a leader in library and information science and information science fields, with its library and information science program ranking first in the United States since 1996. Against the backdrop of "Double First-Class" discipline construction, we hope to provide references for data science curriculum construction and talent training in Chinese universities through

investigating UIUC iSchool's data science course offerings.

2 Related Research

Literature review reveals that foreign scholars primarily focus on theoretical research of data science as a discipline, practical research on data science education and teaching, and applied research on data science. In terms of disciplinary theory, renowned Danish computer scientist and Turing Award winner P. Naur first formally proposed the term "Data Science" in 1974. In the preface to his monograph *Concise Survey of Computer Methods*, he elaborated on the connotation of data science and distinguished it from "Datalogy" [4]. Subsequently, D. Conway proposed the Data Science Venn Diagram in 2010, establishing data science's disciplinary position at the intersection of machine learning, mathematics and statistics, and domain expertise [5].

In data science education and teaching practice research, foreign scholars have focused on specific undergraduate data science courses at representative institutions. For example, P. Anderson et al. [6] described the data science curriculum plan and implementation experience at the College of Charleston in South Carolina, USA. B. Baumer and S. College [7] introduced the data science teaching modules at Smith College, one of the prestigious Seven Sisters colleges, including data visualization, data manipulation/data wrangling, computational statistics, machine learning (or statistical learning), and extended topics such as spatial analysis, text mining, data exploration, and network science, providing feasible suggestions for cultivating students' data processing abilities. R. Veaux and M. Agarwal [8] detailed the 2016 summer undergraduate data science curriculum guidelines at the Park City Mathematics Institute (PCMI), aiming to provide structural references for colleges planning data science programs. V. Song and Y. Zhu [9] proposed a layered Data Science Education Framework composed of three pillars of data science (people, technology, and data), computational thinking, data-driven paradigms, and the data science lifecycle. Based on this framework, they implemented user-based, tool-based, and application-based data science courses at Drexel University. Additionally, foreign scholars have elaborated on the overall relationship between iSchools and data science education, with V. Song and Y. Zhu [10] arguing that iSchools are central hubs for data science education, where interdisciplinary faculty teams in library and information science can cultivate numerous data scientists with diverse skills and broad perspectives. In applied data science research, foreign studies frequently involve government governance, business management, healthcare, life sciences, finance and economics, and data journalism.

In recent years, domestic scholars have begun investigating data science programs, focusing on comparative analyses of domestic and international data science teaching practices. Regarding curriculum construction, Chao Lemen and Yang Canjun et al. [11] surveyed the global status of data science curriculum construction, summarizing common characteristics, consensus experiences, and challenges. Subsequently, Chao Lemen and Xing Chunxiao et al. [12] in-

vestigated the characteristics of data science programs at eight world-class universities from the perspective of distinctive courses. Li Shasha, Zhou Jingwen et al. [13] comparatively analyzed data science and big data-related majors at 14 domestic and international universities at both undergraduate and graduate levels, providing suggestions for constructing big data talent training models.

Regarding iSchools' data science education programs, Yan Hui and Zhang Yuhao et al. [14] investigated 141 courses related to data science education across 10 iSchools alliance institutions, classifying them into 12 categories: foundational theory, related disciplinary theory, statistics, machine learning, data visualization, data analysis, data science tools, data mining, database and data management, social impact of data, data policy and regulation, and self-directed learning. Su Rina et al. [15] studied 15 iSchools offering data science graduate programs from perspectives including disciplinary advantages, system division, curriculum objectives, core courses, and credit systems, exploring issues in library and information science (LIS) discipline construction and talent training under data science. Deng Shengli and Fu Shaoxiong [16] investigated the LIS discipline construction at Drexel University's College of Computing and Information from educational systems, academic research, and social practice dimensions, exploring new developments in LIS under the integration of data-driven, computational science, and information science.

In summary, domestic research on iSchools' data science education remains in its infancy, primarily employing website investigations and focusing on macro-level comparative analyses of data science education programs and curriculum construction. These studies lack in-depth exploration and analysis of micro-level details of data science courses at top-tier foreign institutions, lack field research on data science courses, and lack multi-dimensional investigations of teaching details. As an international leader in library and information science and the benchmarking institution for Zhengzhou University School of Information Management's undergraduate education innovation project, UIUC iSchool serves as the focus of this study. Therefore, this paper combines our institution's actual situation and domestic educational practices to thoroughly investigate the data science curriculum group at our benchmarking institution through field study supplemented by website investigation, obtaining first-hand materials to support data science curriculum construction and development. We conduct micro-level analysis from four dimensions—target student types, teaching formats, degree of instructional collaboration, and course content—to provide references for our institution and domestic data science curriculum construction.

3 Data Sources and Research Methods

This study's data sources include field investigation and website research. From July 2018 to July 2019, one of the authors, Yang Ruixian, visited UIUC and completed four data science courses: *Data, Statistical Models, and Information*; *Information Organization & Access*; *Foundations of Data Science*; and *Theory and Practice Data Cleaning*. She participated in all courses and discussions

through a combination of online and offline formats. In October 2019, both authors searched the UIUC iSchool website (https://ischool.illinois.edu/) using "data" as the keyword, selected "Type=Course," and obtained 42 course entries as the research subjects for data organization and analysis.

During the investigation, we closely integrated classroom practice insights, course-shared materials, and website course information, employing statistical analysis, comparative analysis, and inductive summarization to systematically organize and deeply analyze details of 42 data science-related courses at UIUC iSchool, focusing on course names and descriptions, credit hours, teaching formats, target audiences, and faculty information.

4 Overview of UIUC Data Science Curriculum Group

Different scholars define data science differently. J. Stanton [17] considers data science an emerging field related to collecting, preparing, analyzing, visualizing, managing, and preserving large-scale data. V. Dhar [18] views data science as the study of obtaining knowledge from data. F. Provost and T. Fawcett [19] define data science as the principles, processes, and techniques for understanding existing phenomena through automated data analysis. The University of Michigan Data Science Initiative (DSI) [20] considers data science a series of processes connecting scientific discovery with practice, involving the collection, management, processing, analysis, visualization, and interpretation of large-scale heterogeneous data often related to transformative, interdisciplinary scientific applications.

Through inductive analysis of various definitions, it is evident that data science research and application objects are large-scale data batches, with basic processes including data collection, organization, processing, and presentation. In 2015, Stanford University Statistics Professor D. L. Donoho [20] explicitly stated in his report on 50 years of data science that complete data science can be divided into six major components: data exploration and preparation, data representation and transformation, data computation, data modeling, data visualization and presentation, and data science-related sciences, as shown in Figure 1 [Figure 1: see original paper].

Based on this framework, we classified and analyzed UIUC iSchool's data science courses. Table 1 shows the specific situation. Data exploration and preparation accounts for approximately 16.67%, with core courses including *Foundations of Data Science*, *Foundations of Data Curation*, and *Foundations of Information Processing*, emphasizing both theory and practice in database management and architecture while strengthening systematic thinking. Data representation and transformation also accounts for 16.67%, with core courses such as *Theory and Practice of Data Cleaning* and *Metadata in Theory and Practice*, focusing on students' programming abilities to process data using computers. Data visualization and presentation accounts for 9.52%, with core courses including *Data Visualization* and *Data Science Storytelling*, emphasizing communication

of information and knowledge hidden in data. Data computation accounts for 4.76%, focusing on *Introduction to Cloud Computing* and *Advanced Topics in Machine Learning & Social Computing*. Data modeling accounts for 7.14%, with the core course being *Data, Statistical Models, and Information*. The "Data Science + Discipline" category has the highest proportion at approximately 45.24%, mostly electives such as *Competitive Intelligence and Knowledge Management*, *Scientific Data Policy Seminar*, *Practical Health Data Analytics*, *Data Ethics*, *Bioinformatics Problems and Research*, *Social Media Analytics*, *Predictive Analysis in Finance*, etc., demonstrating the school's broad curriculum coverage, novel teaching content, strong student choice flexibility, and benefits for deepening the cultivation of various specialized talents, enabling students to gain abilities to cope with employment pressures through personalized course customization.

UIUC iSchool employs diverse and personalized teaching formats with three academic terms annually: Spring, Fall, and Summer. Spring and Fall terms have nearly identical course offerings with only minor instructor adjustments and format changes. Among the investigated data science curriculum group, Summer term courses include *Business Analytics*, *Foundations of Information Processing*, *Introduction to Databases*, *Metadata in Theory and Practice*, and *Copyright for Information Professionals*. Three of these five Summer courses—*Business Analytics*, *Foundations of Information Processing*, and *Metadata in Theory and Practice*—are also offered in Spring and Fall, while *Introduction to Databases* and *Copyright for Information Professionals* are only offered in Spring, indicating high-quality, popular courses with strong student satisfaction. Summer courses total approximately 24-48 credit hours, shorter than Spring/Fall courses (32-64 hours). Efficiently completing key and popular courses within a shorter timeframe meets student learning needs and significantly increases these courses' impact. Additionally, the university's international short-term exchange program—Global Education and Training (GET)—is a Summer term feature. For example, in July 2018, undergraduate students from top Chinese universities funded by GET studied the Network Analysis course (a data visualization and presentation category course in the data science curriculum group) taught by Associate Professor J. Diesner at UIUC's School of Information Sciences.

5 Analysis of UIUC Data Science Curriculum Group Settings

Based on the overall classification, quantity distribution, and characteristic terms of UIUC iSchool's data science courses, further exploration of UIUC's data science education and teaching activity experiences is needed. Specifically, we analyzed from four dimensions: target student levels, teaching formats, faculty collaboration, and course content.

5.1 Analysis of Target Student Levels

Our statistics show that 91% of UIUC iSchool's data science courses target master's students, 7% target undergraduates, and 2% target doctoral students, indicating a relatively mature master's training system (see Figure 2 [Figure 2: see original paper]). At the undergraduate level, due to insufficient mathematical and statistical foundations and limited professional knowledge accumulation, many data science courses cannot follow the data science process and schedule for teaching plans. Therefore, UIUC iSchool's data science curriculum group only includes three undergraduate courses: *Introduction to Data Science*, *Foundations of Information Processing*, and *Database Design and Prototyping*. These courses do not require prior programming background but provide students with solid database theory foundations and methods for solving abstract problems using programming languages, preparing them for applications in data analysis, data science, text mining, digital libraries, and knowledge management. UIUC's School of Information Sciences also lacks detailed training objectives and plans specifically for data science degrees at the doctoral level. The only existing doctoral-level data science course, *Advanced Topics in Machine Learning & Social Computing*, is open to all doctoral students on campus. Course content primarily involves deep learning, generative adversarial networks, adversarial learning, word embeddings, and selected hot topics in artificial intelligence (especially bias in data learning, data fairness, and data ethics). Doctoral-level data science courses adopt seminar formats where students deeply discuss papers on these topics, analyze them within broader theoretical, methodological, and domain contexts, and reflect on them within their own research backgrounds.

5.2 Analysis of Teaching Formats

UIUC employs three teaching formats: on-campus courses, online courses, and blended courses (on-campus & online). UIUC uses the Moodle learning management system to facilitate independent course selection, teaching information release, course material and assignment upload/download, group discussions, and academic report presentations. Before 2019, Moodle utilized the Blackboard Collaborate Ultra web conferencing system for weekly real-time synchronous sessions, providing two-way audio/video, whiteboards, breakout rooms, and screen sharing. Starting Spring 2020, the web conferencing system will gradually transition to Zoom.

In UIUC's School of Information Sciences data science curriculum group, on-campus courses account for 55.1%; blended courses account for 34.5%; and online courses account for 10.3% (see Table 2). This indicates that most data science courses still adopt traditional on-campus formats, such as *Database Administration and Scaling for IS* and *Metadata in Theory and Practice*. Blended courses partially employ innovative real-time synchronous methods, allowing students to independently choose learning time, space, and format, such as *Theory and Practice Data Cleaning*. Pure online courses constitute a smaller proportion, mostly concentrated in the "Data Science + Discipline" category,

such as *Competitive Intelligence and Knowledge Management*. In the era of the fourth research paradigm—data-intensive scientific discovery—relying solely on traditional on-campus formats often cannot meet scientific research activity needs. In the big data era, when the target audience is graduate students, new blended teaching models better meet faculty and student needs for improving classroom quality and enhancing data literacy.

5.3 Analysis of Faculty Collaboration

Independent instruction refers to courses taught by a single instructor (preparation, teaching, assessment, etc.), while collaborative instruction involves two or more faculty members and teaching assistants. Using these definitions, we analyzed UIUC iSchool's data science courses, obtaining the faculty collaboration analysis shown in Figure 3 [Figure 3: see original paper]. Overall, independent instruction courses are more numerous (16 courses), while collaborative instruction courses are fewer (13 courses). Most independent instruction courses concentrate in the "Data Science + Discipline" category (8 courses). Except for data exploration and preparation courses (all collaborative) and data computation courses (all independent), other categories show relatively balanced distribution between independent and collaborative instruction.

Figure 3 reveals distinct collaboration patterns across the six categories. Excluding courses with incomplete information, data exploration and preparation includes three collaborative courses, all using on-campus formats with two instructors each. *Data Mining* and *Foundations of Information Processing* employ blended on-campus & online instruction. Data representation and transformation includes three independent and three collaborative courses. Two independent courses—*Open Data Mashups* and *Theory and Practice Data Cleaning*—use real-time synchronous innovative methods, where classroom activities are updated in real-time to designated sections of the teaching system, allowing students to independently choose learning time, space, and format, providing more possibilities for different student types. Data computation courses are all independently taught using on-campus formats. *Introduction to Cloud Computing* focuses on various cloud service application scenarios, covering key concepts such as public, private, and hybrid clouds, APIs, and data security. *Advanced Topics in Machine Learning & Social Computing* engages students actively in deep learning, generative adversarial networks, adversarial learning, and seminars on selected AI topics, requiring instructors to quickly gauge student comprehension. Therefore, UIUC's School of Information Sciences adopts on-campus formats for data computation courses, aligning with students' knowledge absorption, problem digestion, and reflective understanding patterns, demonstrating the school's meticulous student-centered teaching design and educational philosophy.

Data modeling includes one independent and two collaborative courses. *Data, Statistical Models, and Information*—one of only three required courses for UIUC's Master of Science in Information Management—involves statistics, machine learning, R language, and practical applications with heavy and diverse

teaching tasks and enormous information volume. Traditional independent instruction would compromise teaching quality and student engagement. Therefore, this course is collaboratively taught by Associate Professor V. Torvik and Adjunct Lecturer J. Naiman with three teaching assistants (Chengyue Jiao, Xiaoliang Jiang, and Pingjing Yang). The course topics are listed in Table 3 , covering data model information overview, R language data analysis introduction, probability and Bayesian theorem, random variables, expectation and variance, data inference foundations, numerical vs. categorical data, linear models, multiple linear regression, classification, and logistic regression. The course first reviews probability theory, analyzes common probability distributions as information modeling tools, then introduces parametric and non-parametric prediction models and their extensions in unsupervised learning. Throughout, the course emphasizes model selection, quality measurement, and applications of statistical probability models in information management tasks (e.g., prediction, ranking, and data reduction). The instructors' research directions are shown in Table 4 . Vetle Torvik focuses on mathematical optimization, computational statistics, text and data mining, and literature-based discovery, while Jill Naiman specializes in efficient and engaging data visualization methods in sciences. Their collaborative instruction leverages each instructor's expertise, enhancing the quality of this required master's course.

Data visualization and presentation includes two independent courses (*Advanced Data Visualization* and *Network Analysis*) and two collaborative courses (*Data Visualization* and *Data Science Storytelling*). Foundational courses are mostly collaboratively taught, while advanced courses in specialized fields are more effectively taught independently by domain experts. The "Data Science + Discipline" category shows higher independent instruction rates, as these courses focus on specific problem research and exploration, are highly practical, cover broad disciplines, and have smaller student audiences than required or foundational courses, such as *Data Ethics*, *Practical Health Data Analytics*, *Business Analytics*, and *Digital Humanities*.

5.4 Course Content Analysis

We conducted Chinese word segmentation and preliminary keyword frequency statistics on all course descriptions in the data science curriculum group, removing numerals, adverbs, prepositions, conjunctions, and other semantically meaningless words, retaining only nouns, verbs, and proper nouns. Based on keyword weight indicators (Score), we identified the top fifteen keywords (see Table 5). The keyword weight indicator (Score) is primarily determined by keyword frequency, IDF (inverse document frequency), and semantic aggregation with other words in the text [21]. Analysis reveals that UIUC's data science courses extensively cover technology, visualization, conceptual introduction, modeling and practical application, data analysis, structuring, exploratory learning, metadata, social media, and data mining. For example, *Data Visualization* covers the history of data visualization and modern technologies applied to quantitative,

statistical, and network-centric datasets.

To further explore programming languages, methodological tools, and technical application domains in UIUC's data science curriculum, we manually selected nouns, foreign characters, and proper vocabulary related to tools, methods, programming languages, and skills from the 42 course descriptions. We imported these keywords into word cloud software for intuitive visualization, producing Figure 4 [Figure 4: see original paper]. Word clouds visualize keyword importance through font size, color, or thickness, enabling readers to quickly grasp key information. Figure 4 shows that UIUC's School of Information Sciences emphasizes teaching and practicing skills including methods, tools, programming languages, and software, primarily Python, R, and SQL (Structured Query Language). For instance, *Foundations of Data Science* first teaches working in Unix command prompts, then introduces Python programming, focusing on aspects relevant to data science and specific Python modules. Python is primarily introduced and used through IPython or Jupyter notebooks, covering Numpy, Scipy, Matplotlib, Pandas, Seaborn, and Scikit-learn modules, demonstrated through simple data science tasks (data acquisition, cleaning, visualization, and basic analysis). *Business Analytics* uses tools including R, MySQL, and Tableau. Database courses require students to master writing basic queries in SQL and comprehensively understand relational database theory. Students learn machine learning techniques, including supervised and unsupervised learning, dimensionality reduction, and clustering, with emphasis on practical applications to high-dimensional numerical data, time series data, image data, and text data. Finally, students learn to use relational databases and cloud computing software components such as Hadoop, Spark, and NoSQL data stores.

UIUC's data science curriculum also emphasizes source code management software like git and GitHub. These skill-based knowledge areas involve broad application domains including privacy, communication, business, law, academic research, libraries, healthcare, policy standards, ethics, literacy, community, and geography. Combined with Table 2, the "Data Science + Discipline" category's highest proportion is validated here.

Based on job requirements for data science positions and interviews with renowned data scientists, commonly used tools include [22]: (1) Data science languages like R, Python, Haskell, Clojure, Scala; (2) NoSQL tools like MongoDB, Couchbase, Cassandra; (3) Traditional databases and data warehouses like SQL, DW, RDBMS, OLAP; (4) Big data computing tools like Hadoop HDFS & MapReduce, Spark; (5) Big data management, storage, and query tools like Pig, HBase, Hive, Cascalog, Impala; (6) Data collection, aggregation, and transmission tools like Web scraper, Avro, Flume, Hume, Sqoop; (7) Data mining tools like Pandas, SciPy, Weka, Knime; (8) Visualization tools like Tableau, Gephi, Shiny, D3.js, ggplot2; (9) Statistical analysis tools like SPSS, Matlab, SAS. UIUC's data science curriculum construction keeps pace with market demand changes, covering extensive domains and focusing on data exploration and preparation, data transformation processing, data computation

and modeling tools, and data skills construction for data analysis and visualization, providing robust support for comprehensively training professional data talents and optimizing data science education.

6 Conclusions and Recommendations

As a founding member of the iSchools core leadership group (iCaucus), UIUC has long been committed to advancing library and information science education. Facing opportunities and challenges from the fourth research paradigm, UIUC has begun exploring data science-related courses. Through field study and website investigation, we found that UIUC's data science curriculum can be divided into six categories, serving undergraduate, master's, and doctoral students through blended online-offline teaching methods, with collaborative faculty instruction and content that closely follows data science job market requirements.

China's current data science education system and curriculum are still in their infancy, with considerable room for improvement. Based on our investigation of UIUC's data science curriculum construction and comparative analysis from four dimensions (target student types, teaching formats, faculty collaboration, and course content), we propose the following four recommendations for China's data science education:

(1) **Strengthen continuity in data science education** to cover all stages from undergraduate to doctoral. Current data science education in most Chinese universities concentrates at the undergraduate level, with preliminary master's training and scarce doctoral training. As described in Section 5.1, UIUC has a relatively mature master's curriculum system, uses seminars at the doctoral level, and continuously improves undergraduate courses that provide foundational database theory for core specialized courses. Therefore, Chinese universities should aim to provide systematic data science education and curriculum covering all degree levels.

(2) **Enrich innovation in data science teaching** to further integrate blended teaching methods. As described in Section 5.2, UIUC's School of Information Sciences primarily uses the multifunctional Moodle system, transitioning to Zoom in 2020 with enhanced features. China already has quality open online courses on MOOC platforms, such as Professor Chao Lemen's *Introduction to Data Science* at Renmin University and Professor Song Tian's *Python Data Analysis and Visualization* at Beijing Institute of Technology. Using MOOCs to improve teaching, developing quality online courses, and integrating learning management platforms into daily teaching activities remain directions for continuous improvement in Chinese universities.

(3) **Enhance teaching collaboration among data science faculty** to deepen cooperative instruction. As described in Section 5.3, all data exploration and preparation courses at UIUC use collaborative instruction,

and faculty come from diverse backgrounds including traditional LIS, computer science, and library data management practice. Chinese universities have begun inviting librarians into classrooms for collaborative teaching, but inter-faculty collaboration within departments needs strengthening. Faculty recruitment should also adopt interdisciplinary perspectives, hiring compound talents based on practical needs and increasing the number of adjunct and visiting professors.

(4) **Improve research directions in data science curricula** to align content closely with market demands. Most domestic data science courses have computer science and statistics backgrounds, with few courses dominated by library data practice, medical informatics analysis, or similar fields. Data science courses are crucial for enhancing data literacy, cultivating data thinking, and developing high-skilled knowledge professionals—data librarians—in the big data era [23]. As described in Section 5.4, UIUC's data science curriculum covers extensive domains, closely follows LIS application practices and data skill construction, and provides robust support for training professional data talents. Therefore, Chinese universities should thoroughly investigate domestic information market job requirements and selectively integrate practical tools and languages into course experiments and hands-on operations, referencing excellent course content and designs from top foreign institutions.

References

[1] Ministry of Industry and Information Technology of the People's Republic of China. Big Data Industry Development Plan (2016-2020) [EB/OL]. [2019-10-21]. http://www.miit.gov.cn/n1146295/n1652858/n1652930/n3757016/c5464999/content.htm.

[2] TalkingData. Professional Data Talent Education Industry Ecosystem Report [R/OL]. [2019-08-30]. http://mi.talkingdata.com/report-detail.html?id=2.

[3] Ministry of Education of the People's Republic of China. Deeply Promoting "New Engineering" Construction [EB/OL]. [2019-11-27]. http://www.moe.gov.cn/jyb_{xwfb}/xw_{fbh}/moe

[4] NAUR P. Concise Survey of Computer Methods [M]. Lund, Sweden: Studentlitteratur, 1974: 1-30.

[5] CONWAY D. The Data Science Venn Diagram [EB/OL]. [2019-08-31]. http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram.

[6] ANDERSON P, BOWRING J. An Undergraduate Degree in Data Science: Curriculum and a Decade of Implementation Experience [EB/OL]. [2019-10-21]. https://blogs.valpo.edu/data-desk/files/2017/04/Charleston-SIGCSE14-design-of-Data-Science-with-Objectives.pdf.

[7] BAUMER B, COLLEGE S. A Data Science Course for Undergraduates: Thinking with Data [J]. The American Statistician, 2015, 69(4): 334-342.

[8] VEAUX R, AGARWAL M. Curriculum Guidelines for Undergraduate Programs in Data Science [EB/OL]. [2019-10-21]. https://www.stat.berkeley.edu/~nolan/Papers/Data.Science.Gui

[9] SONG V, ZHU Y. Big Data and Data Science: Opportunities and Challenges of iSchools [J]. Journal of Data and Information Science, 2017, 2(3): 1-18.

[10] SONG V, ZHU Y. Big Data and Data Science: What Should We Teach? [J]. Expert Systems, 2016, 33(4): 364-373.

[11] Chao Lemen, Yang Canjun, Wang Shengjie, et al. Empirical Analysis of Global Data Science Curriculum Construction Status [J]. Data Analysis and Knowledge Discovery, 2017, 1(6): 12-21.

[12] Chao Lemen, Xing Chunxiao, Wang Yuqing. Research on Characteristic Courses for Data Science and Big Data Technology Majors [J]. Computer Science, 2018, 3(45): 3-10.

[13] Li Shasha, Zhou Jingwen, Tang Jintao, et al. Analysis of Data Science and Big Data Talent Curriculum System [J]. Computer Engineering and Science, 2018, (40): 109-113.

[14] Yan Hui, Zhang Yuhao, Zhang Xincan, et al. Investigation on Data Science Education Programs in iSchools Alliance [J]. Information and Documentation Services, 2018(4): 95-100.

[15] Su Rina, Yang Qin. Research on Data Science Curriculum System in LIS Discipline: Centered on iSchools Universities [J]. Library Tribune, 2019, 39(4): 40-49.

[16] Deng Shengli, Fu Shaoxiong. LIS Discipline Construction Under the Integration of Computational Science and Information Science [J]. Information and Documentation Services, 2019, 40(3): 80-87.

[17] SONG V, ZHU Y. Big Data and Data Science: Opportunities and Challenges of iSchools [J]. Journal of Data and Information Science, 2017, 2(3): 1-18.

[18] DHAR V. Data Science and Prediction [J]. Communications of the ACM, 2013, 56(12): 64-73.

[19] PROVOST F, FAWCETT T. Data Science and Its Relationship to Big Data and Data-Driven Decision Making [J]. Big Data, 2013, 1(1): 51-59.

[20] DONOHO D. 50 Years of Data Science—A Presentation at the Tukey Centennial Workshop [R/OL]. [2020-05-30]. http://www.mathsccnu.com/forum.php?mod=attachment&aid=MzQ

[21] Official Authority Release: "Tuyue" Hot Word Analysis Indicator Explanation [EB/OL]. [2019-11-24]. https://mp.weixin.qq.com/s?___{biz}=MjM5MDAzOTk4OA==&mid=20419092

[22] Chao Lemen. Data Science [M]. Beijing: Tsinghua University Press, 2016: 31-32.

[23] Wei Lai, Gao Xiran. Role Positioning of University Data Librarians Under Big Data Background [J]. Information and Documentation Services, 2015(5):

90-94.

Author Contributions: Yang Ruixian: Conceptualization, data collection, review and revision; Wan Jiaqi: Data collection, organization, analysis, writing and revision.

Note: Figure translations are in progress. See original paper for figures.