ISBN: 978-981-14-2064-1

Curriculum Design for Students in Mathematics: Data Science Courses as Examples

Chia Hung Kao

Department of Applied Mathematics, National Taitung University
No. 369, Sec. 2, University Road
Taitung, Taiwan
chkao@nttu.edu.tw

Abstract

The STEM (Science, Technology, Engineering, and Mathematics) academic discipline has attracted a lot of attention in recent years. Many universities begin to embrace the trends and propose several innovations and refinements in education, including teaching improvement, integrated curriculum, multidisciplinary programs, and capstone courses to better incubate students for future. In this paper, the experience of curriculum design and implementation for data science courses to the undergraduate program in mathematics is introduced. Based on the original focus on mathematical and computation theories, the employment of data science courses can provide students with the comprehensive theoretical and practical knowledge and skills for the challenges and industrial requirements nowadays. In addition, the curriculum is designed and implemented from the perspectives of students who major in applied mathematics (non-CS students) to better stimulate the learning, participation, exercise, and innovation. Major findings and lessons learned from the refined courses are discussed for more insight into the following deployment and refinement of the curriculum.

Key words: Curriculum design, Mathematics, Data science, Undergraduate program

Introduction

The STEM (Science, Technology, Engineering, and Mathematics) academic discipline has attracted a lot of attention in recent years [1]. In addition, the emerging trends in cloud computing, big data analytics, IoT, financial technology, and artificial intelligence have strong impacts on higher education [2]. Many universities begin to embrace the trends and propose several innovations and refinements in education, including teaching improvement [3], integrated curriculum, multidisciplinary programs [4], and capstone courses [5] to better incubate students for future [6].

Several studies analyzed the traditional mathematics curricula, investigated hidden issues in current practice, and proposed refinements to improve mathematics education. For instance, Coffland and Xie [7] pointed out that there existed three major issues about traditional mathematics curricula, including (1) the gap between course materials and students' "real-life" mathematical experiences [8], (2) the scattering of knowledge in different courses [9], and (3) the lack of mathematics applications to other subjects [10]. For curricula refinements and education improvements, Coffland and Xie also suggested the following characteristics in futhre math curriculum, including (1) the connection between mathematics course content and real life, (2) the connection among related topics, and (3) the connection between mathematics to other subjects as an interdisciplinary approach [11].

Based on the aforementioned challenges and suggestions, a series of course reviews was performed in the department of applied mathematics in one university to get more insight into the major issues and the curriculum improvements possibilities. For instance, students enrolled in statistics course usually learned the theoretical knowledge and did related practices in textbooks by topics. However, the scenarios described in textbooks or course materials are often limited in a certain scope or designed in the simpler form. A critical gap exists between theoretical knowledge and practical problem-solving in the real world. In addition, students have no experience in the leverage of knowledge and skills learned from mathematics, statistics, computer science, and data science which are essential for research and industry nowadays [12].

According to the course reviews and the education requirements, the curricula of certain courses in the department of applied mathematics are refined and implemented. Major courses selected and discussed in this paper are "Data Processing and Analysis" and "Statistical Software for Data Analysis." The education goals, syllabuses, and the corresponding course designs are introduced. Furthermore, major findings and lessons learned are discussed for more insight into the deployment and refinement of the curriculum in the future.

Curriculum Design and Implementation

A. The Design of Data Processing and Analysis Course

The "Data Processing and Analysis" course helps students to realize and experience every important step of data processing and analysis tasks by the leverage of knowledge and skills in mathematics, statistics, computer science, and data science.

In the beginning, as described in Table I, the steps of data processing and analysis, including data collection, data cleaning, data management, data modeling, data analysis, and data visualization, are introduced for students to develop the global picture and the overall understanding. Students can get more insight into the roles of mathematics, statistics, computer science, and data science in the process of data processing and analysis. Furthermore, open data and big data are introduced for students to realize real-world trends, challenges, and problems they will encounter nowadays.

TABLE I
DATA PROCESSING AND ANALYSIS

Topic	Description

Course Introduction	Introduction to Data Processing and Analysis
The Process of Data Analysis	Major steps in data processing and analysis
Open Data	Introduction to open data
Python Basis	Python Review
Data Collection and Management	Using NumPy and Pandas packages for data collection and management
Data Analysis and Visualization	Using statistics library and Matplotlib package for data analysis and visualization
Network and Graph	Using NetworkX package for network and graph modeling
Text Mining	Using Beautiful Soup, lxml, and jieba packages for text mining from social networks
Image Processing	Using OpenCV package for image processing and recognition
Machine Learning	Using Scikit-Learn package for classification, clustering, and regression
Advanced Topics or Case Studies	Selected topics in machine learning or artificial intelligence

After the preliminary introduction, the course introduces each step of data processing and analysis along with related computer science and mathematics knowledge. More specifically, students can learn and experience how to perform data processing and analysis through Python and related packages based on the knowledge and skills accumulated in computer science and mathematics-related courses. In addition, open data will be used as major materials for learning and practice throughout the course. Popular formats of open data, including CSV, JSON and XML, and the corresponding manipulation methods will be introduced for students to perform data collection, data cleaning, data management, data analysis, and data visualization tasks [13]. Necessary Python packages, including NumPy, Pandas, SciPy, SymPy, and Matplotlib, are also introduced as fundamental tools for students on data science and scientific computation [14].

During the course, students are also encouraged to do their assignments and group projects based on open data and to innovate according to their interests and creativities. When students are equipped with solid data processing and analysis knowledge and skills, advanced applications such as network and graph modeling, text mining, image processing and recognition will be introduced. Furthermore, emerging topics and associated frameworks for machine learning and artificial intelligence can be introduced. Based on the knowledge and skills of Python and related packages, students can learn and use frameworks like Scikit-Learn, Tensorflow, PyTorch, and Keras efficiently for particular machine learning problems based on their interests. It is expected that the understanding of global view on data science, the capability of integration on mathematics and computer science, and the learning of programming skills for advanced data processing and analysis tasks can be developed in the "Data Processing and Analysis" course.

TABLE II
STATISTICAL SOFTWARE FOR DATA ANALYSIS

Topic	**Description**
R Introduction	Introduction and installation of R
The Process of Data Analysis	Major steps in data processing and analysis
R Basis	Variables, vectors, and fundamental operations
Data Structures	Using matrix, array, list and data frame
Open Data	Introduction to open data
Data Collection and Management	Data management and manipulation in R
Flow Control	Using if statements, switch and loops
Data Visualization	Using R for data visualization
Statistics	Using R for descriptive and predictive statistics
Regression	Using R for regression analysis
Data Mining	Using R for clustering, classification, and association rule
Case Studies	Financial and social network mining by R

B. The Design of Statistical Software for Data Analysis Course

The goal of "Statistical Software for Data Analysis" course helps students to learn and experience the overall process of data processing and analysis using R programming language. R programming language provides efficient facilities for data manipulation, calculation, statistics analysis, and visualization. Both R programming language and Python are popular tools for data processing and analysis tasks. Therefore, it is worthwhile for students in the department of applied mathematics to learn and practice those skills for various data processing and analysis requirements. Based on previous knowledge of statistics, Python, and the process of data processing and analysis, students can acquire additional capabilities when performing data processing and analysis tasks in the course. Table II describes the syllabus of the "Statistical Software for Data Analysis" course in the department of applied mathematics.

Similar to the "Data Processing and Analysis" course, the steps of data processing and analysis are introduced first for students to develop the overall understanding. Open data will also be introduced and used as course materials for students to learn and experience the data collection, cleaning, management, modeling, analysis, and visualization tasks by R programming language. Similarly, students are encouraged to do their assignments and group projects based on open data and to innovate according to their interests and creativities. In addition to exploratory data analysis, descriptive statistics, and predictive statistics, selected topics in data mining, including clustering, classification, and association rule, are introduced for students. Furthermore, students can get more insight into the applications of data analysis through case studies on financial mining, text mining, and social network mining. It is expected that the insight of global view on data science, the capability of integration on mathematics and computer science, and the learning of R programming skills for data processing and analysis tasks can be developed in the "Statistical Software for Data Analysis" course.

Major Findings

A. The Connection between Course Content and Real Life

One of the issues about traditional mathematics curricula is the gap between course materials and students' "real-life" mathematical experiences. Therefore, both in "Data Processing and Analysis" and "Statistical Software for Data Analysis" courses, in addition to practice statistical theories or data analysis methods through limited or simplified examples in textbooks, open data can be served as great materials for learning [15]. Through the introduction of open data in the courses, students can find and use interesting open data for further analysis, application development, and even innovation. Based on statistical theories and programming skills learned before, students can learn and practice how to explore, collect, clean, manage, analyze, and visualize actual data from open data portal. Therefore, students can not only realize the data processing and analysis methods but also get insight into the applications of statistical theories and programming skills in real world.

Fig. 1 and Fig. 2 present the assignments implemented by Python and R based on government open data. The first one is to process, analyze, and visualize the dengue fever cases reported in major cities in Taiwan by Python, NumPy, Pandas, and Folium in "Data Processing and Analysis" course. The second one is to process, analyze, and present the location of wi-fi hotspots in Tainan city in Taiwan by R, ggplot2, ggmap, and googleVis in "Statistical Software for Data Analysis" course. Through the connection between course content and real-life application based on open data, students can realize the global picture of data processing and analysis, and experience major steps in data processing and analysis comprehensively, including data collection, data cleaning, data management, data modeling, and data visualization rather than limited or simplified examples in textbooks. In addition, complexities and problems during the processing and analysis tasks can help students to develop solid knowledge and skills which are useful for future research and career.

Fig. 1 Data Visualization of Dengue Fever Cases.

Fig. 2 Visualization of Wi-Fi Hotspots.

Furthermore, it is discovered that students show great enthusiasm for the learning and the practice due to the use of interesting open data. Besides, innovative applications of data processing and analysis can be found in the assignments and group projects during the courses. According to these positive results, the connection between course content and real-life is successfully established for students. In the future, more connections can be further carefully identified, designed, and deployed to all the mathematics, data science, and computer science courses for students.

B. Interdisciplinary Learning

Another issue about traditional mathematics curricula is the lack of "the connection between mathematics to other subjects as an interdisciplinary approach." In order to facilitate the learning, practice, and knowledge accumulation in an interdisciplinary approach, lots of assignments and group projects based on mathematics, statistics, computer science, and data science are designed. Students are encouraged to integrate what they learn in different courses and to innovate based on their interests and creativities.

Fig. 3 presents the project built by students during the "Data Processing and Analysis" course. Students integrated what they learned in front-end development, computer programming, data processing and analysis, statistics, and data visualization methods to build a responsive web page for their travel in Japan. The project also got an award in the competition of college of science and engineering in the university. Based on the feedback from the student, the student said it was a wonderful experience to integrate what she learned in the past and develop a great website from scratch. Through the course and the project, she had more confidence and was not afraid of computer-related tasks.

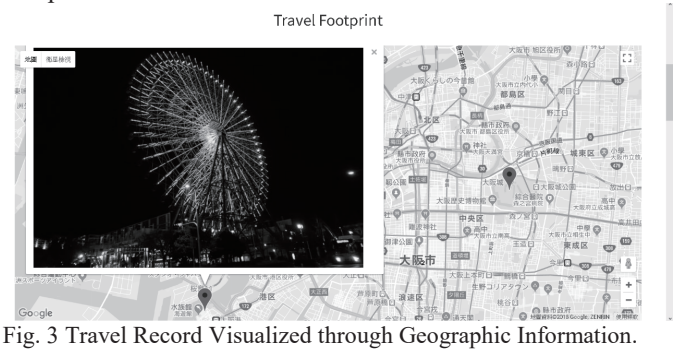

Fig. 3 Travel Record Visualized through Geographic Information.

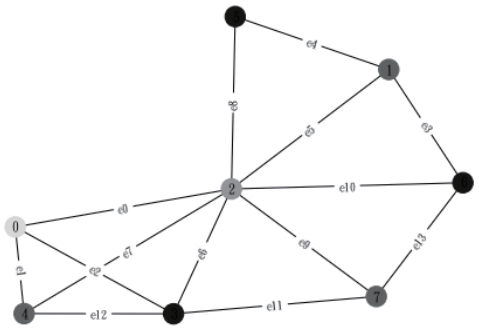

Fig. 4 Graph Coloring by Python and NetworkX.

Fig. 4 presents another project developed during the "Data Processing and Analysis" course. Student who also enrolled in "Graph Theory" course tried to model and implement a graph coloring algorithm to present the coloring result. Although student encountered many difficulties, he solved those problems and completed the implementation by searching, reading, and trying lots of online resources and code examples.

Based on the feedback from those students, in addition to interdisciplinary knowledge, hands-on experience, and problem-solving skills, they got great pleasure and confidence after the project. It is believed that through the interdisciplinary learning, students can learn, practice, and accumulate various knowledge and skills in different disciplines efficiently [16]. Besides, confidence, innovation, and creativities of students can be incubated during the learning activities.

Conclusions

Many universities begin to embrace the trends and propose several innovations and refinements in education, including teaching improvement, integrated curriculum, multidisciplinary programs, and capstone courses to better incubate students for future. In this paper, the curriculum design and implementation for data science to the undergraduate program in mathematics are introduced. Based on the original focus on mathematical and computation theories, the employment of data science courses can provide students with the comprehensive theoretical and practical knowledge and skills for the challenges and industrial requirements nowadays.

Based on the course evaluations, the deployment of data science courses and the refined curriculum bring lots of benefits and positive impacts on the students in the department of applied mathematics. Furthermore, major findings, course experiences, and lesson learned from the course implementation, including "the connection between course content and real life" and "interdisciplinary learning," are found to be highly beneficial to the learning activities. In addition, hands-on experience, problem-solving skills, confidence, innovation, and creativities of students can be incubated during the courses effectively and efficiently. Through the findings and lesson learned, better curriculum can be developed to incubate students for better future.

References

[1] A. Asghar, R. Ellington, E. Rice, F. Johnson, and G. M. Prime, Supporting STEM Education in Secondary Science Contexts, *Interdisciplinary Journal of Problem-Based Learning*, 6(2), 2012, pp. 85-125.

[2] Y. Kim, H.-E. Chu, and G. Lim, Science Curriculum Changes and STEM Education in East Asia, in M. S. Khine (eds), *Science Education in East Asia: Pedagogical Innovations and Research-Informed Practices*, Springer, Cham, 2015, pp. 149-226.

[3] M. Farren, Y. Crotty, T. Lonergan, S. Soutirou, and W. McTiernan, EU Action in Developments in the Curricula and Teaching Methods Appropriate for the Future Global Technical Challenges. *2014 Information Technology Based Higher Education and Training*, York, UK, Sept. 2014, pp. 1-7.

[4] J. Williams, W.-M. Roth, D. Swanson, B. Doig, S. Groves, M. Omuvwie, R. B. Ferri, and N. Mousoulides, *Interdisciplinary Mathematics Education: A State of the Art*, Springer International Publishing, 2016.

[5] J. Wilbarger and S. Howe, Current Practices in Engineering Capstone Education: Further Results from a 2005 Nationwide Survey, *Proceedings of the 36th Annual Conference on Frontiers in Education*, San Diego, CA, USA, Oct. 2006, pp. 5-10.

[6] J. C. Adams and R. J. Pruim, Computing for STEM majors: enhancing non CS majors' computing skills, Proceedings of the 43rd ACM technical symposium on Computer Science Education, Raleigh, North Carolina, USA, Feb.-Mar., 2012, pp. 457-462.

[7] D. A. Coffland and Y. Xie, The 21st Century Mathematics Curriculum: A Technology Enhanced Experience, in X. Ge, D. Ifenthaler, and J. M. Spector (eds), *Emerging Technologies for STEAM Education*, Springer International Publishing, 2015, pp. 311-329.

[8] J. Boaler, *What's Math Got to Do with It? Helping Children Learn to Love Their Least Favorite Subject - and Why It's Important for America*, Penguin, New York, 2008.

[9] G. Wiggins and J. McTighe, *Understanding by Design, Association for Supervision and Curriculum Development*, 2005.

[10] Z. Usiskin, The Current State of the School Mathematics Curriculum, in B. Reys, R. E. Reys, and R. N. Rubenstein (eds), *Mathematics Curriculum: Issues, Trends, and Future Directions*, National Council of Teachers of Mathematics, 2010, pp. 25–40.

[11] K. Weber, M. Inglis, and J. P. Mejia-Ramos, How Mathematicians Obtain Conviction: Implications for Mathematics Instruction and Research on Epistemic Cognition. *Educational Psychologist*, 49(1), 2014, pp. 36–58.

[12] P. B. Henderson, W. Barker, S. Epp, and W. Marion, Math Educators, Computer Science Educators: Working Together, *The 34th SIGCSE Technical Symposium on Computer Science Education*, Reno, Nevada, USA, Feb. 2003, pp. 236-237.

[13] Y. Zhu, Introducing Google Chart Tools and Google Maps API in Data Visualization Courses, *IEEE Computer Graphics and Applications*, 32(6), 2012, pp. 6-9.

[14] J. Bernard, Use Python for Scientific Computing, *Linux Journal*, 2008(175), 2008.

[15] D. Mouromtsev and M. d'Aquin, *Open Data for Education: Linked, Shared, and Reusable Data for Teaching and Learning*, Springer International Publishing, 2016.

[16] P. Charlton and K. Avramides, Knowledge Construction in Computer Science and Engineering When Learning Through Making, *IEEE Transactions on Learning Technologies*, 9(4), 2016, pp. 379-390.