Real Data and Application based Data Science Education in Engineering

Q. Peter He[1], Jin Wang[1], Shiwen Mao[2], Laura Parson[3],
Bo Liu[4], Peng Zeng[5], Allen Smith[6], Daniel Henry[3]

[1]Dept. of Chemical Engineering
[2]Dept. of Electric and Computer Engineering
[3]Dept. of Educational Foundations, Leadership, and Technology
[4]Dept. of Computer Science and Software Engineering
[5]Dept. of Mathematics and Statistics
Auburn University, Auburn, AL 36849
[6]Dept. of Chemical Engineering
Tuskegee University, Tuskegee, AL 36088

Abstract

The democratization of data is transforming our world. Together with the advances in computer and engineering technology, these advancements drive the rapid change in the landscape of jobs and work. There are many reports indicating that industry finds itself constrained by today's relatively small supply of well-trained data science talent, and hiring demand for data scientists has begun to increase rapidly; some projections forecast that approximately 2.7 million new data science positions will be available by 2020. Unsurprisingly, the data science and engineering (DSE) programs across the nation have grown significantly in the past a few years. DSE education requires both appropriate classwork and hands-on experience with real data and real applications. While significant progress has been made in the former, one key aspect that yet to be addressed is hands-on experience incorporating real-world applications. In this work, we will review the efforts that explore real data and application based data science education.

Keywords

Data science, workforce development, project-based education, experiential learning, real-world application

Introduction

Data science and engineering (DSE) is emerging as a field that is revolutionizing the world. There are many interpretations for DSE. All of them center on the notion of multidisciplinary and interdisciplinary approaches to extracting knowledge or insights from large quantities of complex data for use in a broad range of applications[1]. As noted in the summary of a 2018 report "Data Science for Undergraduates: Opportunities and Options" from the National Academies of Sciences, Engineering and Medicine (NASEM), "The continued transformation of work requires both a larger population with a basic understanding of data science and a substantial cadre of talented graduates with highly developed data science skills and knowledge, acquired through substantial coursework and practice." A recent study by IBM found more than 2.3 million data science and analytics job listings in 2015, and predicted that demand for data scientists will soar 28% by 2020 [2]. Unsurprisingly, the data science and engineering (DSE) programs across the nation have grown significantly in the past a few years. As concluded in the NASEM report, DSE education requires both appropriate classwork and hands-on experience with real data and real applications. While significant progress has been made in the former, one key aspect that yet to be addressed is hands-on experience incorporating real-world applications. Specifically, it is insufficient for them to be handed a “canned” data set and be told to analyze it using the methods that they are studying. Such an approach will not prepare them to solve more realistic and complex problems, especially those involving large, unstructured data. Instead, students need repeated practice with the entire DSE cycle beginning with ill-posed questions and “messy” data[1]. In other words, textbooks and traditional lecture courses may offer limited help in developing students’ capability in applying the theory and methods to solve real, complex problems. In this work, we review and summarize the current status of efforts that utilize real data and application in DSE education. Among these efforts, there are live or real project based DSE courses, real data based DSE courses, and other unconventional DSE courses. In the following sections, we review all of them and discuss their advantages and limitations.

Live or real project based DSE courses

In 2016, Saltz and Heckman from Syracuse University reported a case study of a project-focused introduction to big data science course[3]. The course as a case study demonstrated that using live clients within a team-based, project-focused course provides a useful platform in which to teach an introduction to data science course to graduate students across a range of backgrounds. The results of this study indicate that one successful approach is a project-focused class that puts students at the boundary between the academic context of the course and solving a real-world problem for their client. The course, Applied Data Science, was an introduction to graduate students the fundamentals of data science. Two student teams were assigned to two clients. Each client provided data for a real problem, as well as one or more domain experts to help explain the data, the problem to be solved and the business context. Many positive findings have been obtained from this course. For example, in student surveys, 100% of the students agreed with the statement “this course stimulated critical thinking.” In addition, 100% of the students also agreed that the course “provided new viewpoints” of insight. Finally, 92% of the students felt that the course “provided an intellectual challenge.” These findings suggest that the students were engaged and used higher level thinking to work through their challenges[3]. The faculty observed that students voluntarily spent more time on this course, as compared to a traditional/standard class. The project increased student motivation during the class as well as their interest in the field. The authors noted that one of the challenges of teaching a real-world, project-based course is finding organizations willing to participate and share their data and knowledge.

Grisham, Krasner and Perry from the University of Texas at Austin reported a case study of teaching Data Engineering at a graduate-level class using a real-world project[4]. The primary deliverable of the course was a semester-long project to implement an information system in a real-world application domain. The authors believe that the use of such project domains motivate students to apply good Software Engineering principles in the classroom, which consequently encourages those principles to be extended into industrial practice[4]. A similar approach has been reported by Chase, Oakes and Ramsey, where the Small Project Support Center at Radford University has provided the live projects[5]. For both cases, the real-world project came from on-campus centers. Again, one of the challenges in such an approach is finding real-world projects.

It is also clear that the real-world project increased time, organizational, and pedagogical demands on the instructors.

Other similar efforts include Sabin from University of New Hampshire[6]. The common advantages of this type of live project based DSE courses include improving students' motivation and engagement; promoting teamwork; and improving students' written and oral communications skills. The limitations of using live project in DSE education include increased time, organizational, and pedagogical demands, and other burdens on instructor[5,7]; solicitation of live projects is challenging[5]; among others[5,8,9].

Real data based DSE course

Baumer reported a data science course for undergraduates offered through the Statistical & Data Sciences Program at Smith College[10]. 500 million tweets were collected and 500,000 of them were analyzed. The course emphasizes modern, practical, and useful skills that cover the full data analysis spectrum, from asking an interesting question to acquiring, managing, manipulating, processing, querying, analyzing, and visualizing data, as well communicating findings in written, graphical, and oral forms[10].

Gould from University of California - Los Angeles reported DataFest as an undergraduate competition in which student teams have just 48 hours to find and communicate meaning in a rich, complex data set[11]. DataFest had been expanded to include participants from fifteen U.S. colleges and universities as of 2014, provides an opportunity for students to challenge themselves with realistic, large data sets in an intense, fun, and encouraging environment.

Depending on the data source and domain expert support availability, these real data based DSE courses provide similar benefits of those of live or real project based DSE courses but probably to lesser extent. These benefits include improving students' motivation and engagement; promoting teamwork; and improve students' written and oral communications skills. Some of the limitations of real data based DSE courses include: difficult to find assignments that motivate all students[12], and some data sets may not have immediate applications.

Other unconventional DSE courses

Other strategies, such as problem-based learning[13,14], learning by creating[15], have also been proposed in DSE education. But because they do not necessarily use real project or real data, they are not reviewed in this work.

Conclusion

In conclusion, real project or real data based DSE courses offer many benefits to students, including improved motivation and engagement, teamwork, and communications skills. They generally put more burdens on instructors, including time, organizational, and pedagogical demands. In addition, solicitation of live projects can be challenging. Another drawback of these approaches is that although the course model can be adopted by other researchers, there is no learning materials generated that can be widely adopted for enhancing DSE education at other institutions.

References

1. National Academies of Sciences, Engineering, and Medicine. *Data science for undergraduates: opportunities and options*. (National Academies Press, 2018).
2. Columbus, L. IBM predicts demand for data scientists will soar 28% by 2020. *IBM White Pap.* (2017).
3. Saltz, J. & Heckman, R. Big Data science education: A case study of a project-focused introductory course. *Themes Sci. Technol. Educ.* **8**, 85–94 (2016).
4. Grisham, P. S., Krasner, H. & Perry, D. E. Data engineering education with real-world projects. *ACM SIGCSE Bull.* **38**, 64–68 (2006).
5. Chase, J. D., Oakes, E. & Ramsey, S. Using live projects without pain: the development of the small project support center at Radford University. *ACM SIGCSE Bull.* **39**, 469–473 (2007).
6. Sabin, M. A collaborative and experiential learning model powered by real-world projects. in *Proceedings of the 9th ACM SIGITE conference on Information technology education* 157–164 (ACM, 2008).
7. Bickerstaff Jr, D. D. The evolution of a project oriented course in software development. *ACM SIGCSE Bull.* **17**, 13–22 (1985).
8. Farkas, D. Choosing group projects for advanced systems courses. in *ACM SIGCSE Bulletin* **20**, 109–115 (ACM, 1988).
9. Joel, W. J. Realistic student projects. *ACM SIGCSE Bull.* **19**, 244–247 (1987).
10. Baumer, B. A data science course for undergraduates: Thinking with data. *Am. Stat.* **69**, 334–342 (2015).
11. Gould, R. Datafest: Celebrating data in the data deluge. in *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics* (2014).
12. Hardin, J. *et al.* Data science in statistics curricula: Preparing students to “think with data”. *Am. Stat.* **69**, 343–353 (2015).
13. Wlodarczyk, T. W. & Hacker, T. J. Problem-based learning approach to a course in data intensive systems. in *2014 IEEE 6th International Conference on Cloud Computing Technology and Science* 942–948 (IEEE, 2014).
14. Ben-Ari, M. Constructivism in computer science education. *J. Comput. Math. Sci. Teach.* **20**, 45–73 (2001).
15. Doucet, K. & Zhang, J. Learning cluster computing by creating a Raspberry Pi cluster. in *Proceedings of the SouthEast Conference* 191–194 (ACM, 2017).

Q. Peter He

Dr. Q. Peter He is Associate Professor in the Department of Chemical Engineering at Auburn University. He obtained his BS degree from Tsinghua University (China) in 1996, and MS and PhD degrees in 2002 and 2005 from the University of Texas, Austin, all in chemical engineering. His current research interests are in the areas of systems engineering enhanced big data analytics with applications in smart manufacturing, renewable energy, digital agriculture, and cancer and healthcare related research. Besides research, Dr. He is interested in engineering education. His current interest in this area is data science education in engineering.

Jin Wang

Dr. Jin Wang is Walt and Virginia Woltosz Endowed Professor in the Department of Chemical Engineering at Auburn University. She obtained her BS and PhD degrees in chemical engineering (specialized in biochemical engineering) from Tsinghua University in 1994, and 1999 respectively. She then obtained a PhD degree (specialized in control engineering) from the University of Texas at Austin in 2004. The central theme of her research is to apply systems engineering principles and techniques to understand, predict and control complex dynamic systems, including both engineered systems and microbial organisms. Dr. Wang also devotes herself to education and promoting women in engineering.

Shiwen Mao

Dr. Shiwen Mao received a PhD in electrical and computer engineering from Polytechnic University, Brooklyn, N.Y., in 2004. He is the Samuel Ginn Professor and Director of Wireless Engineering Research and Education Center at Auburn University, Auburn, AL. His research interests include wireless networks, multimedia communications, and smart grid. He is a recipient of the Auburn University Creative Research & Scholarship Award in 2018, the NSF CAREER Award in 2010, several conference best paper awards, and The 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He is a Fellow of the IEEE.

Laura Parson

Dr. Laura Parson is an Assistant Professor in the Higher Education Administration Program at Auburn University and Higher Education Administration MEd/PhD Program Coordinator. Her Ph.D. is in Teaching & Learning, Higher Education from the University of North Dakota. Her research interests focus on identifying the institutional practices, processes, and discourses that coordinate the teaching and learning experiences of women in higher education, explored through a critical lens. She is a qualitative methodologist, with a focus on ethnographic and discourse methods of inquiry. Her research questions seek to understand how pedagogy, classroom climate, institutional environment, and curriculum inform student experiences, and how the institution coordinates those factors through translocal practices.

Peng Zeng

Dr. Peng Zeng is currently Associate Professor of Statistics in Department of Mathematics and Statistics, Auburn University. He got a bachelor’s degree in Mathematics from Nankai

University, and a master's and doctorate degree in Statistics in Purdue University, West Lafayette. His main research interests include high-dimensional data analysis, design and analysis of experiments, semi-parameter regression, machine learning, deep learning, etc.

Allen Smith

Dr. Allen Smith is an assistant professor of chemical engineering at Tuskegee University. He received B. S. and Ph.D. degrees from Auburn University and a M.S. from the University of Washington. He worked in the pulp and paper industry for 14 years after the M.S. before returning to school for the PhD. His research interests include Pulp and Paper, Energy from Biomass, Food Science, and Undergraduate education. His teaching responsibilities include Thermodynamics, Reaction Engineering, and Process Control.

Daniel Henry

Dr. Daniel Henry is the founding director of the Auburn Center for Evaluation (ACE). ACE has conducted evaluations for the Pew Charitable Trusts, USDA, McGraw-Hill, and the Alabama State Department of Education. Dr. Henry has a Ph.D. in Educational Psychology from Indiana University, and is an Assistant Clinical Professor at Auburn University. He has published in the areas of program evaluation, qualitative research, educational pedagogy, and hunger studies. He teaches program evaluation, qualitative research, and learning and cognition at Auburn. Before coming to Auburn, Dr. Henry was a faculty member at Central Michigan University and Indiana University, and taught high school English for 13 years.