Infusing Industry Interests into an Applied Data Science Program

Enrique Barbieri, Xuqing (Jason) Wu

Information & Logistics Technology Department
College of Technology
University of Houston

Abstract

Interest on big data, data analytics, and data science has grown in every area of the global economy that increasingly lives in a word built on data. Universities across the globe are reacting by launching or expanding data-related programs from courses to certificates, academic tracks, and complete BS, MS, and PhD degrees. Data science can reveal insights and draw intelligent conclusions on behaviors, preferences, trends, and forecasts by mining the data captured through massive digital transactions. Industry is several steps ahead in realizing the importance of data science toward business competitiveness. However, despite the enthusiasm for this fast-approaching digital future, there is a gap between the business challenges and academic programs offering data science advances. The purpose of this session at the CIEC 2020 is to provide a forum for the exchange of information between higher education and industry partners. Particularly, we offer an opportunity for industry and academia to review the current landscape of classes and programs, identify industry priorities, and propose innovative activities that enable a sustainable ecosystem for a multidisciplinary education program.

1. Introduction

The top 26 paying jobs listed in 2018 by Forbes included data scientist and data architect [1]. In the 2019 list, data scientist moved to #1. Members of industrial advisory boards across university programs express a pressing need for graduates knowledgeable in the data science field. There is on-going debate to define terms such as data science, data analytics, big data, data engineering, and others as these permeate society's buzzwords. The quick definition is that data science is the umbrella field, while analytics, engineering, management, and others are either tracks or subfields or job titles. As it often happens, graduates of non-STEM fields learn data concepts on the go and on the job by attending workshops and earning certificates, while others seek formal academic training via academic degrees. Universities tend to react to these emergent needs by mobilizing academic departments to launch courses, minors, degree programs and certificates.

Even if Data Science emerges as an independent field of study, its foundations are highly interdisciplinary and the data scientist finds a myriad of application fields. STEM fields such as computer science and engineering are used to dealing with data. Perhaps what has changed is the data size, the frequency of accumulation, the rich format variety (numbers, images, videos, text), and the lack of structure. The desirability of exposing students to real data is well recognized [2]. However, industry may be reluctant to provide real data sets because privacy and security concerns, for example in automatic detection of underground structures from seismic data,

extreme climate effects on the housing market, oil well production time-series data, real-estate value estimates, medical doctor classification, computer storage performance, and others. Non-STEM fields also accumulate and use data. For example, political campaign managers use demographic and voting data to deploy strategies. The subfields of data collection, formatting, storage, and retrieval, as well as information extraction and visualization, and data-driven decision-making have taken a front-row seat in universities and in industries alike.

This paper points to various challenges that academic departments face in addressing this emergent field and offers an opportunity for faculty and industry to discuss the relevancy of academic or textbook topics in the context of the application of data science to solve problems laid out by industry. To that end, the session presents four technical articles summarized below which illustrate a range of applied problems that enjoy Data Science solutions, and introduce various areas of research and scholarship at the graduate level. The session also distributes a brief survey designed to gather feedback from industry to assist universities in creating applied data science programs.

The four articles are:

1. "A CNN Based Framework to Classify Anticlines Structures on Seismic Data," by Pablo Rondon, Core Facility for Advanced Computing and Data Science, University of Houston

The paper demonstrates the effectiveness of using machine learning to classify seismic image and recognize key subsurface structures by analyzing seismic data. After a brief instruction to several machine learning techniques that include PCA, KNN, Random Forest, SVM, and CNN, the paper conducts a comparative research with regard to the classification accuracy. The result claims CNN the winner for its superior capability in identifying complex geological structures from seismic data.

2. "Data Science and Traditional Engineering and Technology Programs – How to Improve Operational Excellence," by Vassilios Tzouanas, Department of Computer Science and Engineering Technology, University of Houston – Downtown

The paper gives an overview and highlights the potential of applying data science to improve operational excellence for process industries. The paper lists several examples of how data analysis can assist the process industry to improve productivity and minimize environmental impact. The paper suggests that current engineering curriculum could be enhanced with three to four courses to help undergraduate students master data analysis skills in their domain.

3. "Big Seismic Data Visualization on Cloud: A Case Study Collaborating with Industry," by Hao Zheng and Lei Huang, Department of Computer Science, Prairie View A&M University, and Bidur Bohara, and Mike Heck, Materials and Structural Analysis Division, Thermo Fisher Scientific, Houston, TX.

This paper introduces a cloud-based platform for visualizing big 3D seismic data. The platform is built on top of open-source software packages with a proprietary 3D graphic engine. The paper outlines the technical details in pushing for visualizing large data sets in near real-time. The paper also highlights the benefit for students from the collaboration with the industry.

4. "A Robotic Scout UAV for Mapping Dynamic Environments at Bechtel Corporation," by Ahmed Abdelmoamen Ahmed, Olumide Abel and Adeoluwa Akinwa, Department of Computer Science, Prairie View A & M University

Supported by the industry, this paper presents a design for using a UAV to scan for damages inside pipelines, a challenge faced by maintenance engineers. Due to the loss of GPS signals, the paper proposes an alternative navigation method by using LiDAR and a computational method. The system and the autonomous navigation algorithm are tested in a simulated physical environment. Encouraged by the simulation result, a real system is built for testing.

2. University Programs

The traditional university approach when creating a new program is to rely first on the existing expertise of the faculty in a particular department and a set of existing courses across departments and even across colleges. The program may be simply a set of courses, or a formal minor, or a degree program, or a certificate for non-university professionals. Many individuals sometime overlook the fact that academic courses necessarily have a pre-requisite structure. For example, a course that explores sampling variability and statistical estimates may place as a pre-requisite a course on probability and statistics often taught in the mathematics department. In the university lingo, these are hidden pre-requisites, which lengthen the program and its cost. Unless the course presentation is in an executive style, it is often impossible to cover background knowledge (math, statistics, programming) while also addressing the subject matter in any meaningful way. Sometimes pre-requisites are waived at the expense of course depth to attract a larger audience, shorten the program, and reduce the cost.

New courses are developed either for students majoring in the department or as part of minors offered by that department. The issue of course pre-requisites discussed above is not trivial and affects a department's ability to launch a given program successfully. Departments sometimes elect to set aside three or four data science courses as a cohort of electives giving a student a way to earn a specialization track under her/his traditional degree plan. Complete degrees and certificate programs follow similar constructs. This structure is useful in the Data Science field because it permits students to develop specialized competency within specific disciplines. The downside is that it does not promote data science as an over-arching set of concepts and tools useful and applicable to a broad spectrum of activities. In any case a university's mix of ingredients to produce a data science mindset should include data science foundational knowledge, real-world data and problem-solving skills, and ethical considerations [3].

The following programs at the University of Houston exemplify the model just described:

1. The Computer Science Department is launching this fall 2019 a Data Science Minor for Non-Majors, comprised of 5 courses, four of which are in the Computer Science Department covering programming with Python, three courses on Data Science topics, and a fifth course that can be taken from the student's own major department [4].

2. As a result, the CIS program in the College of Technology has approved the CS Data Science Minor as part of its ABET-accredited BS degree enabling majors to earn the very attractive combination of a BS plus minors in Business Administration and in Data Science.

3. In fall 2017, the Mathematics Department launched a hybrid MS in Statistics and Data Science, focusing on statistical inference, modeling and analysis of complex data. The degree can be completed in one year [5].

4. The HP Enterprise Data Science Institute at the University of Houston (http://www.uh.edu/data-science-institute/)is collecting information about graduate data-related courses offered across campus. The intent is to launch a website where non-university individuals select a set of courses (possibly from various departments) based on particular interests, academic background, and working skills, to earn a graduate certificate.

5. UH Downtown MS in Data Analytics prepares students in advanced statistics, digital data acquisition, digital data management, data analysis, and data presentation [6].

6. The Information and Logistics Technology Department in the College of Technology at UH envisions launching an MS in Applied Data Science that seeks to infuse a high degree of industry interests into the courses. The preliminary plan shown in Table 1 follows the structure described above, that is, a set of existing or new courses based on faculty expertise are organized as core, electives, and project or thesis options. This approach is not unlike what one finds in any other institution. One difference is that we seek to attract industry's feedback into several of the courses so the eventual graduates will be more competitive.

3. Industry Interests

Industry professionals find it challenging and sometimes impossible to discern between a degree program and another, or even one certificate from another, or to navigate through the many similarly named course options across several departments. Typical questions they may ask are Is a whole semester course on programming needed? How much statistics do I need? How do I meet the course pre-requisites? Should I pursue analytics or science? What is the difference between the Introduction to Data Science course in the Computer Science Department and the one in the Mathematics Department?

On the other hand, and of interest to academics, is what questions industry is interested in answering. That is, what practical problems is industry facing that prompts their interest in an academic program in the field of Data Science? It is important to find out so that courses reflect those interests and attempt to provide the tools needed to answer their questions.

The second goal of this paper is to bring industry to the design stage of a set of courses in data science. To that end, we offer the survey instrument shown in Table 2.

Table 1. Draft plan for an MS in Applied Data Science

MS APPLIED DATA SCIENCE 30 Credits (Draft)		
COURSES	**EXISTING**	**NEW**
CORE 18 Credits		

TEPM 6301 - Project Management for Technical Professionals	X	
CIS 6321 - Principles of Information Systems Security	X	
CIS 6311 - Programming for Data Science		X
CIS 6312 - Data Science	X	
CIS 6313 - Cloud Computing		X
CIS 6314 - Information Visualization		X
THESIS/PROJECT OPTIONS 12 Credits		
Project: 3 credit project + 9 credit electives		
Thesis: 6 credit thesis + 6 hours electives		
ELECTIVES 6 Credits (with Thesis) or 9 Credits (with Project)		
CIS 6315 - Big data analytics		X
ELET 6397 - Biomedical Data Mining	X	
ELET 6397 - MATLAB for Engineering Technology	X	
ELET 6397 - Applied Statistics for Technology	X	

Table 2. Industry survey to assess topics of interest for an Applied Data Science program

1: Strongly Agree 2: Agree 3: Disagree 4: Strongly Disagree 5: N/A					
Mark the appropriate box based on the rubric above	**1**	**2**	**3**	**4**	**5**
1. Data Science & Analytics are relevant to my company operations					
2. The following areas are important:					
a. Programming for data analytics (e.g. Python)					
b. Visualization/presentation (e.g. use of Tableau)					
c. Regression analysis: relate two or more variables of interest					
d. Big Data mining: uncovering patterns in large data sets					
e. Image processing					
f. Text processing					
g. Speech processing					
3. Tool utilization in various application fields (e.g., TensorFlow)					
4. One or more members of my team is considering:					
a. Advanced degree in Data Science/Analytics (e.g., MS)					
b. Graduate Certificate program at a University					
c. Undergraduate Certificate program at a University					
5. My company is interested in working with academic programs to provide real data-sets for students to learn from					
Other questions or concerns specific to your company/business operations or contact the authors at ebarbieri@uh.edu:					
Contact Email (optional):					

4. Conclusions

Data Science & Analytics are disruptive areas of interest to industry and academia. Industry struggles with filling its day-to-day operations that increasingly rely on data. Academic struggles with identifying the most efficient pathway for students and professionals to quickly become proficient in data analysis. This paper presented some of the real challenges that Universities face in meeting industry's needs, and their solution approach to create courses, degrees, and certificate programs primarily based on existing courses and faculty expertise. The main goal of this effort is to involve industry professionals in academia's curriculum decision-making process in order to improve the competitiveness of our graduates. A future article will report on the University of Houston's success across the various fronts described in this paper.

References

[1] The 25 Highest-Paying Jobs in America in 2018, Janice Gassam, Senior Contributor, Forbes https://www.forbes.com/sites/janicegassam/2018/08/15/the-25-highest-paying-jobs-in-america-in-2018/#393022005fd5

[2] NASEM, 2018b. Envisioning the data science discipline: The undergraduate perspective: Interim report. Tech. rep., National Academies of Sciences, Engineering, and Medicine, Washington, DC, https://doi.org/10.17226/24886.

[3] NASEM, 2018a. Data science for undergraduates: Opportunities and options. Technical Report, National Academies of Sciences, Engineering, and Medicine, Washington, DC, https://doi.org/10.17226/25104.

[4] University of Houston's Computer Science Department Minor in Data Science for Non-Majors http://www.uh.edu/nsm/computer-science/undergraduate/programs/minor-data-science/

[5] University of Houston's Mathematics Department Minor in Statistics and Data Science http://www.uh.edu/nsm/math/graduate/ms-statistics-datascience-outline/

[6] University of Houston – Downtown MS in Data Analytics https://www.uhd.edu/academics/sciences/pages/master-in-data-analytics.aspx

Biographical Information

ENRIQUE BARBIERI received the PhD in Electrical Engineering from The Ohio State University and was on the faculty of the School of Engineering at Tulane University as an assistant professor of Electrical Engineering (1988-94), associate professor of Electrical Engineering (1994-96), associate professor of Electrical Engineering & Computer Science (1996-2002), and chair of Electrical Engineering & Computer Science (1996-98). During 2002-11, he was on the faculty of the College of Technology at the University of Houston as professor and chair of Engineering Technology (2002-09), associate dean for research and graduate studies (2009-10), member of the Executive Council of the TX Manufacturing Assistance Center (2006-11), chair of the Council (2007-09), and director of the Center for Technology Literacy (2006-10). During 2012-18, he was professor and chair of Engineering Technology, College of Engineering at the University of North Texas. In 2018, he joined the University of Houston's College of Technology as professor and chair of Information & Logistics Technology. He is a senior member of IEEE – Control Systems Society, a member of ASEE, and was an active contributor to the ASEE Engineering Technology Council and the Engineering Technology Leaders Institute (2003-2018).

XUQING (JASON) WU received the PhD in Computer Science from the University of Houston in 2011. He is an assistant professor of Computer Information Systems in the College of Technology at the University of Houston. His research is focusing on developing data analytic methods and exploiting mathematical models suitable for solving real-world problems. Particularly, Dr. Wu is interested in innovative machine learning approaches and fundamentals of predictive analytics. Related research endeavors include subsurface sensing, statistical inversion, computer vision, high performance, mobile and cloud computing.