Bringing Data Science, Xinformatics and Semantic eScience into the Graduate Curriculum

P. Fox

Rensselaer Polytechnic Institute, Tetherless World Constellation, Troy, United States (pfox@cs.rpi.edu)

Recent advances in acquisition techniques quickly provide massive amount of complex data characterized by source heterogeneity, multiple modalities, high volume, high dimensionality, and multiple scales (temporal, spatial, and function). In turn, science and engineering disciplines are rapidly becoming more and more data driven with goals of higher sample throughput, better understanding/modeling of complex systems and their dynamics, and ultimately engineering products for practical applications. However, analyzing libraries of complex data requires managing its complexity and integrating the information and knowledge across multiple scales over different disciplines.

Attention to Data Science is now ubiquitous - The Fourth Paradigm publication, Nature and Science special issues on Data, and explicit emphasis on Data in national and international agency programs, foundations (Keck, Moore) and corporations (IBM, GE, Microsoft, etc.). Surrounding this attention is a proliferation of studies, reports, conferences and workshops on Data, Data Science and workforce. Examples include: "Train a new generation of data scientists, and broaden public understanding" from an EU Expert Group, "...the nation faces a critical need for a competent and creative workforce in science, technology, engineering and mathematics (STEM)...", "We note two possible approaches to addressing the challenge of this transformation: revolutionary (paradigmatic shifts and systemic structural reform) and evolutionary (such as adding data mining courses to computational science education or simply transferring textbook organized content into digital textbooks).", and "The training programs that NSF establishes around such a data infrastructure initiative will create a new generation of data scientists, data curators, and data archivists that is equipped to meet the challenges and jobs of the future."

Further, interim report of the International Council for Science's (ICSU) Strategic Coordinating Committee on Information and Data (SCCID), features this excerpt from section 4.2.4 Data scientists and professionals: "An unfortunate state in the recognition of data science, is that there is a lack of appreciation of the need for a set of professional knowledge in skill in key areas, many of which have not been emphasized to date, e.g. professional approaches to the management of data over its lifecycle. As such, the effort required to be a data scientists is not valued sufficiently by the remainder of the scientific community." SCCID Recommendation 6 reads: "We recommend the development of education at university level in the new and vital field of data science. The curriculum included in appendix D can be used as a starting point for curriculum development. Appendix D. is entitled "Example curriculum for data science" and explicitly uses the "Curriculum for Data Science taught at Rensselaer Polytechnic Institute, USA" .

This contribution will present relevant curriculum offerings at the Rensselaer Polytechnic Institute. http://tw.rpi.edu/web/Courses