Learning Big Data Analytics with Digital Storytelling

Jungwoo Ryoo\textsuperscript{1}, Soo-yong Byun\textsuperscript{2}, Dongwon Lee\textsuperscript{3}, Eun-Kyeong Kim\textsuperscript{4}

\textsuperscript{1}College of Information Sciences and Technology, The Pennsylvania State University, Altoona, PA 16601, USA, \textsuperscript{2}Department of Education Policy Studies, The Pennsylvania State University, University Park, PA 16802, USA, \textsuperscript{3}College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA 16802, USA, \textsuperscript{4}GeoVISTA Center, Department of Geography, The Pennsylvania State University, University Park, PA 16802, USA.

Keywords: big data, big data analytics, big data mining, digital storytelling, learning module, iSchool.

This paper aims at introducing a part of our NSF research project on big data education, “Building a Big Data Analytics Workforce in iSchools\textsuperscript{1}.” There is increasing demand for skilled personnel in big data industries, but existing big data curricula at the university level focus primarily on students with a strong computational background, ignoring a large segment of students who might otherwise pursue education and training in this vital area, but who will be faced with big data issues in the workplace. Therefore, our project’s primary goals are to address the national demand for professionals with knowledge in big data and to broaden the pool for a big data analytics workforce. Part of this effort will involve research as to whether the newly developed learning modules are more effective at increasing students' big data competencies, e.g., knowledge, skills, and analysis.

The objective of this project is to develop three innovative learning modules. These modules will be designed to: (i) utilize both group-based and contextualized learning methods and (ii) be applicable and accessible to students majoring in disciplines outside, but related to main-stream computer science (e.g., iSchools). The first module will involve digital storytelling exercises where students will be asked to develop their own narratives about the relevance and significance of big data in solving real-life problems and will be expected to become knowledgeable of, and proficient with, big data concepts and applications. The second module will be more technical in nature and will allow students to discover the efficacy of big data concepts in solving practical problems in information security. Finally, the third module will introduce more advanced topics in big data mining, such as examining a large amount of complex data to unearth important patterns and knowledge, and introducing how to interpret the results to arrive at appropriate decisions in a specific context.

Module 1 Details

As of this writing, we are finalizing the first phase of our project and completing our first module. The goal of the first module is to use “storytelling” to build awareness about big data, big data analytics techniques, and big data-related career opportunities. Since a majority of iSchool students whom module-1 targets may not have obtained basic computational competencies (i.e., concepts of algorithms, data structures, or programming), in the first module, our intention is to first introduce students to basic concepts on big data and encourage them to explore the materials in the format familiar to them (instead of programming). Digital storytelling is the modern expression of the ancient art of storytelling (Barrett, 2005), and has gotten increasingly popular in recent years, as shown well in the examples of YouTube, podcasting, or Second Life. Today’s young generation is very familiar with diverse platforms and tools (e.g., Second Life, OpenSim, Xtranormal, Garry’s Mod, or Microsoft Kinect) with which interactive

\textsuperscript{1} http://sites.psu.edu/bigdata/
videos can be easily made, played, and shared. Three examples are shown in Figure 1. We have an experience in using Second Life and OpenSim to develop security educational materials, as shown in Figure 1(c) below (i.e., NSF DUE-0817376 TUES phase 2 project--An Immersive Security Education Environment Using Second Life).

(a) Second Life  
(b) Xtranormal  
(c) PIs’ I-SEE Project

Figure 1. Example Virtual Platforms for Digital Storytelling

In this project, we attempt to answer the following research questions: How to develop innovative teaching materials in the big data analytics context to effectively train undergraduate students with insufficient computing competencies? To determine whether our new learning modules are more effective to increase students’ awareness, knowledge, and skills about big data analytics as well as computational competencies, compared to conventional teaching methods. Toward this end, we will employ quasi-experimental designs that use pre-posttests either with or without control groups depending on the availability of control groups. Specifically, we will use a group comparison design with pretests and posttests for students at University Park and Altoona. That is, we create a treatment (i.e., a course taught by the new teaching modules) vs. a control setting (i.e., a course taught by traditional teaching methods).

Our research will have a direct impact on increasing the number and diversity of undergraduate students with computational competencies and big data skills. Initially, our project will positively contribute to training undergraduate students in 4-year iSchools and 2-year community colleges. Eventually, the pedagogical findings and developed teaching materials will have a positive influence on general computing education with the focus on big data analytics in broader scientific and engineering communities throughout the nation. The developed big data analytics materials through this project will be freely available from our project web site (e.g., http://sites.psu.edu/bigdata/).

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1525601.

References:
