Editorial: Special issue on “Big Data in Education”

Renato P. dos Santos
renatopsantos@ulbra.edu.br

PPGECIM- Doctoral Program in Science and Mathematics Education,
ULBRA-Lutheran University of Brazil, Brazil

Diebold is often quoted as being the author of the first academic reference to ‘Big Data’ in his paper *Big Data’ Dynamic Factor Models for Macroeconomic Measurement and Forecasting* (2000). There, he defined it as referring “to the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology” (Diebold, 2000).

This explosion made Big Data a popular (and effective) marketing buzzword a few years ago. In their best-seller book (2013), Mayer-Schönberger & Cukier put Big Data on a ‘revolutionary’ level comparable with the Internet or the printing press, affirming right on the title that it “[…] Will Transform How We Live, Work, and Think”. In his Wired Magazine article, with the quite provocative title *The End of Theory: The Data Deluge Makes The Scientific Method Obsolete*, Anderson went further to affirm that the way Science has worked for hundreds of years “is becoming obsolete”, in the sense that “with enough data, the numbers speak for themselves” (2008).

According to Gartner, however, Big Data has already crossed the Peak of Inflated Expectations and become more than a mere buzzword. Presently, we see mobility, Internet of Things (IoT), and smart machines technologies mutually reinforcing themselves up to the peak (Gartner Group, 2015).

Indeed, while Internet users are currently producing more and more data of various types such as e-mails, social media, search queries, etc., a pervasive legion of ‘things’, each one having its own IP number, gathers data through wearables, connected cars, and IoT and SmartCity sensors at a rate that is going to push the very idea of ‘big’ to higher levels.

In our less grandiose reality, we are undeniably witnessing Big Data techniques promoting innovations leading to better health care, a cleaner environment, safer cities, more effective marketing strategies, and even new ways of doing science. Nevertheless, the recent WikiLeaks and PRISM incidents disclosed the accompanying increasing risks of re-identification of previously anonymized data, of disclosure of individual’s Social Security number, and of leakage of personal identification at cellular and genetic levels.

In any case, just as Big Data is transforming other industries such as insurance, finance, retail, and professional sport, in time, it will also transform education (DiCerbo, Behrens, & Barber, 2014).

One of such ‘transformations’ is undoubtedly the spate of Big-Data-based MOOCs that has been flooding the field of online education. They give institutions the predictive tools they need to design curricula that collect data at every step of each the dozens of hundreds of thousands of students learning processes. They address students’ necessities with customized modules, assignments, feedback and learning trees in the curricula that will promote better and richer outcomes for individual students (Guthrie, 2013).

Turning to Big Data as a course content, Kate Mueller, formerly a student at Syracuse University, said that to look for good Big Data experts or good data analysts only among
computer science people “is missing the boat in terms of people who see the structure underlying things and solve problems” (Dumbill et al., 2013). Along the same lines, Mattmann (2013) claims that natural scientists, too, should become familiar with Big Data.

Considering the above, more than a mere training in computer infrastructures or predictive analysis, we need preparation for the ethical and scientific challenges posed by Big Data to the real world in which they will exercise their professions. Accordingly, this special issue intended to demonstrate the impact and potential for Big Data to improve both school administration and educational activity, address new methodologies and new applications in education, and discuss the challenges and possibilities that such enlarged scale brings for teaching and learning.

In the first paper, considering the fast-increasing need of data scientists, the growing number of data science programs to educate future data scientists, and the lack of research on the best pedagogy to be used to teach such an introduction course, Saltz and Heckman report on a case study of an introductory Big Data Science course to graduate students across a range of backgrounds. This project was grounded in experiential learning and boundary theory, focused on the learning potential created by the boundaries between school and the client’s organization, by using real-world, live projects, in the sense of serving a real customer, with a real problem, and the students working with real data sets, using the popular “R” open source statistical analysis and visualization system. Even if the study showed the need for a few refinements, results were deemed of value to the clients who provided the data, it provided a meaningful set of skills required to do data science across many domains, and the students reported this course challenged them to think outside the textbook, what was attributed to the boundary condition created.

In the second paper, Dalla Vecchia proposes an ‘inversion’ of the traditional Mathematical Modeling (MM) process in classroom scenarios, which is often described in the literature as a sequence of predefined steps, usually starting with a real problem or situation and then proceeding to develop a model. Standing MM as a unique approach to treating massive volumes of data and draw inferences, in his proposal Big Data and Google Correlate (GC) are prospective means, with particular emphasis on the understanding of how knowledge is constructed. Here, the starting point is the actual development of the model, after which it is input into GC, which spots real internet search behaviour situations that best fit to the model, this fitting being gauged by a correlation coefficient provided by GC, and move towards an understanding of the modelled phenomenon by means of cause-and-effect relationships and scientific explanations. This author sees this working with Big Data and GC going beyond the teaching of Mathematics and being linked with what is called digital literacy, understood as the ability to deal with and interpret digital media, an ability so needed in our present technological world.

In the third paper, considering that human brain classifies words in massive and complex dynamic semantic networks that influences speech, behaviour, and choices of interest, which seems to be a primordial factor for learning to occur more effectively, Bübül presents a methodology based on Wordle, a web page-based word cloud application, commonly used as a Big Data visualization tool. It can assist teachers in understanding their students’ primary needs or interests to decide on contexts to be used in the classroom by presenting a person with a letter of the alphabet and asking them to speak or write a word starting with it, with the assumption that the chosen word usually reflects the structure of the person’s brain and is related to contexts relevant to this person. This author understands that this methodology creates a small gateway into the learners’ world/mind that allows researchers and teachers to understand student’s needs and learning styles.
Closing this issue, dos Santos seeks support in the Philosophy and Constructionism literatures to discuss the realm of the concepts of Big Data and its philosophy and the notions of strong and weak ‘emergence’ to introduce the idea crowledge, the unexpected knowledge that weakly emerges from Big Data and IoT analysis of individuals’ digital footprints spontaneously left in the digital universe. Furthermore, this author discusses his vision of learning-with-Big-Data as a promising way to build scientific knowledge by thinking like a scientist, exploring crowledge through the mediation of accessible Big Data application software, making concrete abstract concepts such as ‘physical laws’ and ‘causations’, what reinforces the need for more conceptually-robust approaches to Dig Data.

Drawing together visions and experiences from different educational contexts and cultures around the world, this special edition aspires to contribute to the further development and advancement of the dialogue among these research pioneers of Big Data in Education at international level.

I would like to express here my warm thanks to these research pioneers of Big Data in Education for having accepted the invitation and for their contributions. I am also most grateful to the anonymous reviewers for their comprehensive and constructive comments and suggestions, which contributed vastly to the scientific worth of this special issue. It is my pleasant duty to acknowledge the help willingly given by Dr. Gregory Piatesky in publicizing our Call for Papers in KDnuggets (http://www.kdnuggets.com), a leading site for Business Analytics, Big Data, Data Science, and Data Mining resources. It was a pleasant learning experience for me, and I must gratefully acknowledge my indebtedness to the Editors, Professors Athanassios Jimoyiannis and Tassos A. Mikropoulos, for their trust in my capabilities. The workload was greatly relieved by having worked with friends.

References