Guest Editorial Preface

Special Issue on Dig Data Research and Internet of Things Research:
A New Digital Vision for the Knowledge Society

Ernesto D'Avanzo, University of Salerno, Fisciano, Italy

Leyla Zhuhadar, Western Kentucky University, Bowling Green, KY, USA

Miltiadis D. Lytras, The American College of Greece, Athens, Greece

SHAPING THE DATA DELUGE

Spurious Correlations or Scientific Method for Big Data?

Calude & Longo (2016), probably, could not have made a better choice of employing Dante's[1] verses in order to describe, with eloquent effectiveness and terseness, the dilemma lived by *data scientists*, whose chief task, as it is now very well known, is to model the "deluge" of data that, every day, sink us. The authors, in their epilogue, always paraphrasing the Italian *supreme prophet*[2], claim that "we cannot rely on computational brute-force, but we must pursue scientific commitment and knowledge". The "brute-force" and the "scientific commitment and knowledge" represent the two *philosophies* to approach the Big Data analysis, respectively "the end of theory" and the "scientific method".

The idea behind "the end of theory" philosophy[3] is that "with enough data, the numbers speak for themselves": algorithms, more or less powerful, are able to find *correlations* and/or *regularities* within huge datasets. Altogether, this approach does not ask to "theorize, understand, criticise"; to Anderson "correlation is enough …". In their work, Calude & Longo argue, cleverly, and with methodological rigor, their objections to the positions *à la* Anderson, providing, at the same time, the mathematical reasons to adopt the *scientific method* for Big Data Analysis. *Correlations* and/or *regularities* found in Big Data represent a new toolbox to extend the *scientific method* approach.

The first paper of this special issue, authored by D'Avanzo and Lytras, provides a short overview of the debate on the *scientific method* and the *machine learning* approach to Big Data analysis, highlighting the limits and potential of machine learning as a reference analytical framework for Big Data. The four subsequent works all fall into the common interpretation introduced in this first work. The paper of Pipitone and Pirrone is an attempt to align ontologies and relational databases, using an Hidden Markov Model, in order to estimate the most likely sequence of symbols that describe the structures in the relational schema extracted from the input database. Matsumura, Miura, Komori

and Hiraishi report on an inquiry aiming at modeling the emotions evoked in order to obtain an overview of positive and negative users' attitudes towards the disaster happened (e.g. Great East Japan Earthquake). Experiments carried out suggest suggests that such a kind of approach represents a mature technology to be adopted as *early warning systems*. In their work, Piccialli, Chianese, and Marulli show how cultural sensitivity mining can be performed through a text mining process to infer social users' trends in their affection towards cultural topics and events. The work introduced through the paper shows how Cultural Heritage could benefit from business and competitive intelligence, demonstrating how the employment of novel strategies for engaging more visitors could be beneficial to cultural organizations and operators. With the advent of Big Data, social analysts (political analysts, brand analyst, and so forth) increasingly need of personalized social analysis tools, as those introduced in the paper authored by D'Avanzo and Pilato. These types of frameworks seem able to monitoring *social attitudes*, in order to support decision makers for social forecasting in the light of the *situation awarness* philosophy.

Taken together, the contributions published in this special issue show, empirically, on the battlefield, how takes shape, from a practical point of view, the debate, depicted above, on the approach to be adopted when dealing with Big Data. For each paper, their authors have carved out their own piece of Big Data, offering a possible empirical solution, based on either their own knowledge of the field (i.e. expertise) or by believing in the resolutive power of a more statistical based approach. In both cases, decision making processes and/or hypotheses generated seem to make use of a combination of abductive reasoning that promises to be successful both from a theoretical and practical point of view, in particular to bring context to consumers' traces and making new theories arise (Magnani, 2009).

Ernesto D'Avanzo
Leyla Zhuhadar
Guest Editors
Miltiadis D. Lytras
Editor-in-Chief
IJKSR

REFERENCES

Calude, C. S., & Longo, G. (forthcoming). The Deluge of Spurious Correlations in Big Data. *Foundations of Science*. doi:10.1007/s10699-016-9489-4

Magnani, L. (2009). *Abductive Cognition*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-03631-6

ENDNOTES

[1]   "… you were not born to live like brutes, but to pursue virtue and knowledge". The verses are taken from Inferno of Divine Comedy (Canto XXVI, lines 118-120). The whole verse is: "Consider your origin; you were not born to live like brutes, but to pursue virtue and knowledge"

[2]   "supreme prophet" standing for "supreme poet" is another way to refer to Dante Alighieri

[3]   As known, one of the biggest advocates of this approach is C. Anderson, former Editor-in-Chief of *Wired* that, precisely on the journal he run, published the provocative essay entitled *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.*