

GUEST EDITORS' INTRODUCTION: SPECIAL ISSUE ON BIG DATA INTELLIGENCE AND COMPUTING

Anna KOBUSIŃSKA ^{*}, Ching-Hsien HSU [†]

In the past several years we have observed a tremendous growth of the amount of data that is continuously being produced by diverse sources, including computers, mobile devices, sensors or social networks, and that is represented in a variety of structures. In order to classify these huge data sets that are increasing exponentially in both complexity and volume, the term Big Data has appeared [5]. Big Data has become an ubiquitous in understanding and solving complex problems, spanning different disciplinary fields such as computer science, information management, computational biology, healthcare, social networks, business, telecommunications, and many others [1, 7, 10]. Also, the applications of Big Data have become endless, and they influence and change every part of business and society [2, 8, 9, 12].

To describe the Big Data phenomenon, the five Vs, namely huge Volume, high Velocity, high Variety, low Veracity and high Value are used [6]. Volume refers to the immense amount of incessantly generated data. This increasingly makes data sets too large to store and analyze by using traditional database technology. Velocity is related to the speed at which new data is generated, and at which it moves around. The speed at which data streams emerge is unprecedented and must be dealt with in a timely manner, which is very challenging. Another issue connected with velocity is that data are being generated continuously. Variety refers to the different types and formats of data we can now use. In the past we focused on structured data that neatly fitted into tables or relational databases. But, in fact, 80 percent of the worlds data is now unstructured (including messages, social media conversations, photos, sensor data, video or voice recordings), and cannot easily be stored using hitherto applied technologies. Moreover, very often there is a need to bring together these differed types of data with more traditional, structured data [4]. Veracity depicts the messiness or trustworthiness of the data. With many forms of Big Data, their quality

^{*}Poznań University of Technology, Institute of Computing Science, Poznań, Poland
(email: Anna.Kobusinska@cs.put.poznan.pl)

[†]Chung Hua University, Department of Computer Science and Information Engineering, Taiwan, (email: chh@chu.edu.tw)

and accuracy are less controllable. Also, the volumes often make up for the lack of appropriate data quality. Therefore, efforts should be made to keep data free of noise and abnormality. However, the real quality of Big Data is not in not only in its large volumes, high velocity or variety, but also in the ability to turn data into value and to clearly understand which benefits it brings. Thus, value refers to the amount of suitable knowledge that can be extracted from data.

In the context of above properties, one of the grand challenges of the 21st century is the ability to properly analyze vast and complex data sets, as it offers a whole new way of understanding the world, by discovering patterns and correlations in datasets. Subsequently, users can use this processed information to gain deeper insights and to get business advantages [11, 14]. In order to exploit such huge volumes of data, new techniques and technologies, along with advanced tools of exploratory data analysis, data mining, machine learning and data visualization are needed.

While Big Data can provide extremely useful information, it also presents new challenges associated with storing and efficiently sharing data [3]. Data streams require an increasing processing speeds, they also should be stored and accessed faster and more efficiently than ever before. Thus, crucial are the answers to questions how much will cost data storing, whether the data will be secure, and how long it must be maintained?

Big Data brings together not only large amounts of data but also various data types that in the former years never would have been considered together. The heterogeneity and variable representations of Big Data make the traditional methods used so far to store and manage data insufficient and impracticable. Therefore, there is a need to devise new data models and technologies that can handle them [13].

In this special issue we have compiled a number of technical contributions devoted to recent advances and research directions in the field of Big Data. Presented papers address experimental and theoretical works, along with examples of their application to real life situations. The contributions focus on the wide range of topics, such as data modeling, graph matching, incremental localization, and social web of services.

The paper entitled "Pre-processing and Modeling Tools for BigData" by H. Hashem and D.Ranc, discusses one of the most important challenges in Big Data, namely, efficient data modeling and retrieval. In the first part of the paper authors summarize common data sources, shortly describe widely used data processing engines and non-relational databases used to store Big Data. They also provide a discussion on advantages and disadvantages of different tools and approaches dealing with modeling Big Data. The second part of the paper is devoted to the approach of dynamic combining different tools of data modeling. The use of the proposed automatic programming software — BigData Workbench, allows to select data coming from different sources, and model the relevant data, and analyze it. The authors conducted experimental evaluation of their solution and described the obtained results and their business impact. The proposed approach increases performance of data calculation, and consequently the overall business costs.

In the second paper, entitled "Polynomial time algorithms for variants of graph

matching on partial k -trees", T. Nagoya considers different variants of graph isomorphism, namely the graph isomorphism with restriction and the prex set of graph isomorphism. The first problem helps to decide a mapping between vertices of two graphs under the constraint that some vertices of the first graph are not allowed to be mapped to some vertices of the second one. In turn, the second variant is applicable when a mapping between vertices of two graphs has to be decided under the constraint that some vertices of the first graph have to be mapped according to a given isomorphism to some vertices of the second one. However, as the first approach is NP-complete, and the second one is graph isomorphism GI - complete, in the paper a polynomial time exact algorithms for both variants on partial k -trees with bounded k is introduced. As the image of isomorphisms can be controlled, these problems have more wide application than graph isomorphism. The author has explained the difference between his contribution and the same problem on chordal graphs, to prove the originality of his approach. Also, a formal evaluation of the proposed protocol is provided. Graphs structures often represent relationship between data, and as such they acquire more and more importance in applications processing big sets of data, as well as in social networks. Thus, a comparison of graph structures is an important topic, having a wide application in today's big data world.

The contribution "Incremental localization algorithm based on regularized iteratively reweighted least square" written by Xiaoyong Yan et.al., proposes an improved method of incremental localization in wireless ad-hoc networks, which intends to overcome current limitations caused by cumulative error propagation and collinearity problem. The authors use the iteratively reweighted method in localization process to guide weight selection, and thus reduce the difference between the variance of the previous locating error and the variance of the posterior locating error. They also use the regularization method in calculation to avoid the collinearity problem between nodes that configure location. The simulations results prove that the effect and applicability of the proposed solution excels in both: high accuracy, and high stability. Therefore, the presented protocol is suitable for different deployment environments and has high adaptability.

Referring to the forth paper "How Internet of Things influences human behaviour Building Social Web of Services via Agent-Based Approach", M. Komarov, N. Konovalov and N. Kazantsev introduce the Social Web of Services. In the paper the authors investigate the impact of smart Things on the user behavior patterns, and check to which extent such patterns are changed in the result of the human-thing interactions. The authors have performed several simulation experiments, which they describe in their contribution. In the performed experiments a Smart home benchmark of Smart Thing business initiative was used, and an agent-based modeling simulation was performed. The obtained results clearly show that the Smart Things are able to change the way humans act. As a result, companies can influence consumers behavior by controlling the IoT networks and setting appropriate modes for their smart products. The results obtained in this contribution can significantly foster business process management in the field of Big Data and Internet of Things.

We thank all the authors who submitted manuscripts to this special issue. The publication of this issue would not have been possible without their efforts. We also would like to thank the reviewers for providing reviews and suggestions that were instrumental to increase clarity, and high quality of manuscripts. Last, but not least, we thank the Editor-in-Chief, Prof. Jerzy Stefanowski, and FCDS editors for all of the effort that has gone into producing this special issue.

References

- [1] V. Mayer-Schönberger, and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Amazon, 2014.
- [2] N. Japkowicz, and J. Stefanowski (Eds.), *Big Data Analysis: New Algorithms for a New Society*, Studies in Big Data, 2016.
- [3] Jaroslav Pokorný, How to Store and Process Big Data: Are Today's Databases Sufficient?, *Proceedings of Computer Information Systems and Industrial Management - 13th IFIPTC8 International Conference*, CISIM 2014, Ho Chi Minh City, Vietnam, November 5-7, 2014, pp.631–643.
- [4] J. Abawajy, Comprehensive analysis of big data variety landscape, *International Journal of Parallel, Emergent and Distributed Systems*, 30:5–14, 2015.
- [5] M.A. Beyer, and D. Laney, *The Importance of Big Data: A Definition*, Gartner, 2012, <https://www.gartner.com/doc/2057415>.
- [6] A.D. Duncan, *Focus on the Three Vs of Big Data Analytics: Variability, Veracity and Value*, Gartner, 2014, <https://www.gartner.com/doc/2921417/focus-vs-big-data-analytics>.
- [7] A. McAfee, and F. Brynjolfsson, *Big Data: The Management Revolution*, Harvard Business Review, 2012.
- [8] J. Amudhavel, V. Padmapriya, v. Gowri, K. Lakshmipriya, K. Kumar, and B. Thiagarajan, Perspectives, Motivations and Implications Of Big Data Analytics, *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering Technology (ICARCSET 2015)*, ACM, 2015, pp. 34:1–34:5.
- [9] A. Gandomi, and M. Haider, Beyond the hype, *International Journal of Information Management: The Journal for Information Professional*, ACM, 35, 2, pp.137-144, 2015.
- [10] H. Chen, R.H.L. Chiang, and V. C. Storey, Business intelligence and analytics: from big data to big impact, *Journal MIS Quarterly*, ACM, 36, 4, pp. 1165-1188, 2012.

-
- [11] A. Ribeiro, A. Silva, and A.R. da Silva, Data Modeling and Data Analytics: A Survey from a Big Data Perspective, *Journal of Software Engineering and Applications*, 8, pp. 617-634, 2015.
 - [12] J.H. Abawajy, A.Y. Zomaya, and I. Stojmenovic, Network computing and applications for Big Data analytics, *J. Network and Computer Applications*, 59:361, 2016.
 - [13] F. Xhafa, L. Barolli, A. Barolli, P. Papajorgji (Eds.), Modeling and Processing for Next-Generation Big-Data Technologies With Applications and Case Studies, *Modeling and Optimization in Science and Technologies* Springer, 2015.
 - [14] D. Gil, and I-Y. Song, Modeling and Management of Big Data: Challenges and opportunities, *Future Generation Comp. Syst.*, 63:96–99,2016.