Journal of Universal Computer Science, vol. 21, no. 6 (2015), 754-756
submitted: 13/2/15, accepted: 26/3/15, appeared: 1/6/15 © J.UCS

Intelligent Distributed Processing Methods for Big Data

J.UCS Special Issue

Jason J. Jung
(Chung-Ang University
Seoul, Korea 156-756
j3ung@cau.ac.kr)

David Camacho
(Universidad Autónoma de Madrid
Madrid, Spain
david.camacho@uam.es)

Costin Badica
(University of Craiova
Craiova, Romania
cbadica@software.ucv.ro)

Today, "Big Data" is a new information overloading problem in many different areas. Such areas include health cares (e.g., medical records, bioinformatics), e-sciences (e.g., physics, chemistry, and geology), and social sciences (e.g., politics) [Bizer et al. 2011, Jung 2009]. Thus, as we have various types of feasible data from a number of available sources, it is becoming increasingly more difficult to efficiently process such Big Data. Distributed computing technologies (e.g., Hadoop, Hive and Pig) are strongly related to the "Big Data" issues [Hogarth and Soyer 2015, Jung 2012]. Given a very large scale "Big Data," efficient distributed data processing and management remain a challenge in many research areas, for example, information acquisition and stream processing, as well as data integration [Madden 2012]. Also, the number of diverse information processing system architectures might be involved in these areas. They need to exploit relevant solutions to support a number of intelligent services (e.g., knowledge management and decision making). The aim of this special issue is to bring together researchers and practitioners in areas of distributed computing to share their visions, research achievements and solutions, to resolve the issues on big data processing and to establish worldwide cooperative research and development. This will give an opportunity to push further the discussion upon the potential of knowledge and semantic systems across many communities.

This special issue is devoted to analysis of these "Big data" sources and what is more important to identify the areas where Big data can be applied and provide the knowledge that is not accessible for other types of analysis. Additionally, applications of Big data can be investigated either from static or dynamic perspective. We seek for business and industrial applications of Big data that help to solve real-world problems. The area of Big data analytics bring together researchers and practitioners from different fields and the main goal of this special issue is to provide for these people the opportunity to share their visions, research achievements and solutions as well as to establish worldwide cooperative research and development. At the same time, we want to provide a platform for discussing research topics underlying the concepts of Big data analytics and its applications by inviting members of different communities that share this common interest of investigating social networks.

The first paper in this issue, authored by Ana I. Torre-Bastida et al., proposes an interesting big data analytical functionalities for heterogeneous databases. Particularly, two complementary use cases have been presented to illustrate the potential of using the open data in the business domain. The first represents the creation of an existing and potential customer knowledge base, exploiting social and linked open data based on which any given organization might infer valuable information as a support for decision making. The second focuses on the classification of organizations and enterprises aiming at detecting potential competitors and/or allies via the analysis of the conceptual similarity between their participated projects

The second paper authored by Paloma Cáceres et al. introduces big data processing scheme to understand public bus networks. The proposed process has studied modeling and linking accessibility data by using ontological knowledge.

In the third paper, Quang Dieu Tran and Jai E. Jung presents the software platform for discovering contents and stories in the movies. They claims that it is an important big data sources for digital cultural contents and understanding our society. The system automatically understand the movies by discovering social networks and measuring various social measurements.

The fourth paper by Zbyněk Falt et al. focuses on parallel data processing and parallel streaming systems for big data analytics. One of the key components of these systems is the task scheduler which plans and executes tasks spawned by the application on available CPU cores. The proposed task scheduler combined with the new memory allocator achieve up to speed up on a NUMA system and up to 10% speed up on an older SMP system with respect to the unoptimized versions of the scheduler and allocator.

In the fifth paper, Héctor Allende-Cid et al. focus on distributed regression problem. A new Distributed Regression System is presented, which makes use of a discrete representation of the probability density functions. Neighborhoods of similar datasets are detected by comparing their approximated pdfs. This information supports an ensemble-based approach, and the improvement of a second level unit, as it is the case in stacked generalization.

The sixth paper by Alejandro Zambrano et al. introduces visualization algorithm that improves understandability of run-time production systems. The visualization system has been designed by the Set of Experience Knowledge Structure (SOEKS).

The seventh paper by Ngoc Tu Luong et al. proposes an efficient method to analyze a large scale publication data. Particularly, recommendation system based on the data is built to assist users for collaboration.

This special issue has been achieved by a number of fruitful collaborations. We would like to thank the editor in chief of Journal of Universal Computer Science (JUCS), Christian Gütl, for his kind support and help during the entire process of publication. The special issue has selected 7 high-quality papers out of 17 submissions (about 41% acceptance rate). This was possible thanks to the work of the renowned researchers that provided their anonymous reviews.

Finally, we are most grateful to the authors for their valuable contributions and for their willingness and efforts to improve their papers in accordance with the suggestions and comments from reviewers.

Jason J. Jung, David Camacho, and Costin Badica
(Seoul, Korea, May 4, 2015)

References

[Bizer et al. 2011] Bizer, C., Boncz, P., Brodie, M.L., Erling, O.: The Meaningful Use of Big Data: Four Perspectives - Four Challenges. *SIGMOD Record*, 40(4):56-60, 2011.

[Hogarth and Soyer 2015] Hogarth, R.M., Soyer, E.: Using Simulated Experience to Make Sense of Big Data. *MIT Sloan Management Review*, 56(2):49-54, 2015.

[Jung 2009] Jung, J.J.: Contextualized query sampling to discover semantic resource descriptions on the web. *Information Processing & Management*, 45(2):283–290, 2009.

[Jung 2012] Jung, J.J.: Evolutionary Approach for Semantic-based Query Sampling in Large-scale Information Sources. *Information Sciences*, 182(1):30-39, 2012.

[Madden 2012] Madden, S.: From Databases to Big Data. *IEEE Internet Computing*, 16(3):4-6, 2012.