

Computational Intelligence in Big Data

Big Data Analytics has become increasingly popular, not only in academia, but also in industrial and government applications, which can be attributed to the fact that Big Data Analytics offers huge promises as well as imposes grand challenges in a large number of critical real-world applications such as business intelligence, finance, healthcare, climate science, life science and cyber security. Data driven and data intensive approaches have considered the Fourth Paradigm [1] in scientific discovery and technology innovation in the coming decades.

The main challenges in handling Big Data lie not only in the four V's, namely, huge Volume in amount, high Variety in type, Velocity in terms of real-time requirements, and Variability, i.e., constant changes in data structure and user interpretation, but also in the approach to understanding data. Therefore, Big Data calls for a revolutionary change in research methodology, and in tools to be employed as well.

Computational Intelligence (CI) techniques are expected to provide powerful tools for addressing Big Data challenges, as the main techniques in CI, such as evolutionary computation, neural computation and fuzzy systems are inherently capable of handling various amount of uncertainty, which makes CI techniques well suited for

dealing with Variability and Variety of Big Data. On the other hand, the other two V's, Volume and Velocity may create serious challenges to existing CI techniques. Consequently, new CI techniques need to be developed to efficiently and effectively tackle huge amount of data, and to rapidly respond to changing situations. It should be pointed out, however, that such new techniques will not be developed from scratch; instead, they are based on many on-going research topics scattered in different areas of CI research, e.g., large-scale optimization [2], many-objective optimization [3], learning in non-stationary environments [4], and natural language processing [5–6]. A recent review of the use of evolutionary computation and other meta-heuristics in optimization of biological systems has been reported in [7].

To react to the drastic developments in Big Data, the IEEE Computational Intelligence Society (IEEE CIS) has taken several measures, including setting up the *First IEEE Symposium on Computational Intelligence in Big Data*, which is going to be held within the SSCI 2014 in December 2014 in Florida [8], and revising the name of the Technical Committee on Data Mining to expand its scope to cover Big Data Analytics.

This special issue of the *IEEE Computational Intelligence Magazine* is one in a series of reactions of the IEEE CIS to the boom in Big Data Analytics research that aims to publish selected research

articles presenting most recent advances and discussing challenges and opportunities in the application of CI techniques to Big Data. After a peer-review procedure, two papers have been selected from 12 manuscripts submitted to the Special Issue to be present in this Special Issue, representing the highest quality work on CI techniques in understanding Big Data challenges. The first paper, titled *The Emerging "Big Dimensionality"* by Zhai et al., discusses one of the most important challenges in Big Data, namely, the big dimensionality CI techniques have to handle in Big Data. In the early 1990's, a classification or regression problem having hundreds of features might be termed as high-dimensional. Nowadays, as the paper shows, the dimensionality of the benchmark problems found in the popular machine learning databases has increased to millions or even tens of millions. For such high-dimensional problems, most existing learning methods deem to fail and feature selection becomes indispensable. The paper continues to review the most widely used techniques for feature selection followed by highlights of major challenges in feature selection. Finally, the paper suggests that in a very high-dimensional feature space, correlation between the features can be very sparse, which may provide a promising line of research to follow up.

The second paper, *Computational Intelligence Challenges and Applications on Large-Scale Astronomical Time Series*

Databases authored by Huijse et al., discusses a specific yet important Big Data application where CI techniques can play an essential role, namely time domain astronomy (TDA). The paper indicates that astronomical time series analysis is characterized by not only a huge size, but also highly uneven and noisy samples, and imbalance in different sample classes. As a result, unsupervised, semi-supervised and active learning techniques have been widely adopted for solving different problems in TDA. Then, the Large Synoptic Survey Telescope is presented as an example for discussing Big Data challenges in the context of TDA. Such challenges, as also seen in other typical Big Data problems range from large

dimensionality, noise, heterogeneity to computationally efficiency.

Both papers are of review nature discussing the opportunities and promises of CI techniques along with challenges. They have raised more questions than answers, which is actually the aim of this Special Issue to trigger a wider range of interests in the CI community.

Finally, we would like to thank Prof. Hisao Ishibuchi, the Editor-in-Chief, for giving us the opportunity to guest-edit this special issue. We are equally grateful to all authors who submitted their work to this special issue and reviewers for providing us timely and insightful reviews, without which it would not have been possible for us to produce the Special Issue.

References

- [1] T. Hey, S. Tansley, and K. Tolle, Eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Washington, D.C.: Microsoft Research, Raymond, 2009.
- [2] Z. Yang, K. Tang, and X. Yao, "Large scale evolutionary optimization using cooperative coevolution," *Inform. Sci.*, vol. 178, no. 15, pp. 2985–2999, 2008.
- [3] H. Ishibuchi, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization: A short review," in *Proc. IEEE Congr. Evolutionary Computation*, 2008, pp. 2424–2431.
- [4] R. Polikar and C. Alippi, "Guest editorial: Learning in non-stationary and evolving environments," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 1, pp. 9–11, Jan. 2014.
- [5] E. Cambria, B. White, T. S. Durrani, and N. Howard, "Guest editorial: Computational intelligence for natural language processing," *IEEE Comput. Intell. Mag.*, vol. 9, no. 1, pp. 19–63, Feb. 2014.
- [6] E. Cambria and B. White, "Jumping NLP curves: A review of natural language processing research," *IEEE Comput. Intell. Mag.*, vol. 9, no. 2, pp. 48–57, May 2014.
- [7] S. A. Thomas and Y. Jin, "Reconstructing biological gene regulatory networks: Where optimization meets big data," *Evol. Intell.*, vol. 7, no. 1, pp. 29–47, 2014.
- [8] [Online]. Available: <http://iee-ssci.org/CIBD.html>



Call for Papers for Journal Special Issues

Special Issue on "Neurodynamic Systems for Optimization and Applications"

Journal: *IEEE Transactions on Neural Networks and Learning Systems*

Guest Editors: Zhigang Zeng, Andrzej Cichocki, Long Cheng, Youshen Xia, Xiaolin Hu

Submission Deadline: August 15, 2014

Publication: May/June 2015

Further Information: Long Cheng (chenglong@compsys.ia.ac.cn)

<http://compsys.ia.ac.cn/~chenglong/ndo.pdf>

Special Issue on "New Trends of Learning in Computational Intelligence"

Journal: *IEEE Computational Intelligence Magazine*

Guest Editors: Guang-Bin Huang, Erik Cambria, Kar-Ann Toh, Bernard Widrow, Zongben Xu

Submission Deadline: August 15, 2014

Publication: May 2015

Further Information: Guang-Bin Huang (egbhuang@ntu.edu.sg)

[http://www.ntu.edu.sg/home/egbhuang/pdf/CIM-2015-05\(New-Trends\)-Final.pdf](http://www.ntu.edu.sg/home/egbhuang/pdf/CIM-2015-05(New-Trends)-Final.pdf)