Editorial

Programming Foundations for Scientific Big Data Analytics

Wenbing Zhao ¹, Longxiang Gao,² and Anfeng Liu ³

¹Cleveland State University, Cleveland, OH, USA

²Deakin University, Burwood, NSW, Australia

³Central South University, Changsha, China

Correspondence should be addressed to Wenbing Zhao; w.zhao1@csuohio.edu

Received 7 March 2018; Accepted 7 March 2018; Published 19 April 2018

Copyright © 2018 Wenbing Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Big data analytics is the process of examining large data sets to uncover hidden patterns and previously unknown correlations. Big data analytics has been widely used in businesses to find market trends, customer preferences, and other useful business information. The research community is also beginning to embrace this exciting and powerful technology. Considering the huge amount of data produced in scientific fields such as biology, medicine, physics, and material science, big data analytics can be a powerful means of making new scientific discoveries. Efficient and effective big data analytics requires the development of programming tools and models.

This special issue attracted 20 high quality submissions. After a rigorous review process, 13 papers were accepted in this issue. The research presented in these papers can be roughly categorized into three areas: (1) platform for big data analytics (3 papers), (2) machine learning algorithms for big data (6 papers), and (3) big data analytics for various applications (4 papers).

Platform for Big Data Analytics. B. R. Chang et al. reported their work on how to integrate popular big data platforms such as Hadoop and Spark to perform high performance big data analytics. They focused on the optimization of job scheduling based on computing features to improve system throughput. L. Zhang and J. Gao introduced a novel incremental graph pattern matching algorithm for big graph data. By batching insert operations together by considering matching states, they were able to demonstrate higher efficiency of the proposed algorithm. L. Yang et al. proposed several optimization algorithms based on node compression to help solve the shortest path problem in the context of routing big data.

Machine Learning Algorithms for Big Data. H.-F. Ke et al. proposed a new optimal weight learning machine that is capable of incremental learning while the network grows in terms of the number of hidden nodes. L. Wang et al. presented an adaptive ensemble method for classification with imbalanced data. They rely on self-adaption based on the average Euclidean distance between test data and training data, which is obtained by the k -nearest neighbors algorithm. W. Aziguli et al. introduced a new algorithm designed specifically for text classification, which could be useful for analyzing text-based big data. The algorithm is based on the use of denoising autoencoder and restricted Boltzmann machine, which has the advantage of better performance in antinoise and feature extraction. Z. Yuan et al. proposed a new matching pursuit method to overcome the singularity problem and improve the stability of extreme learning machine (ELM).

X. Fan et al. reported a new hybrid similarity calculation model, which is essential in many machine learning algorithms. The model was designed specifically for recommendation algorithms by addressing the user interest drift issue. This model uses the function fitting to reflect users' rating behaviors and their rating preferences and employs the Random Forest algorithm for the user attribute features. X. Pu et al. proposed a hybrid biogeography-based optimization algorithm for big data analytics. The algorithm is used with a feedforward neural network model called multilayer perceptron.

Big Data Analytics for Various Applications. Y. Zhu et al. reviewed the latest development on big data management in the field geological information services. They proposed a system architecture and outlined requirements for big data management for this application domain. D. Li et al. proposed

a loss aversion cooperation model for behavioral economics in crowd-sensing. They showed that their model encourages higher cooperation rate with lower pay rate.

S. Zhang et al. introduced a new algorithm to make prediction on advertisement click-through rate. The algorithm is based on the weighted extreme learning machine and the Adaboost algorithm. A more accurate predication on click-through rate would increase advertising performance, which could lead to the improvement of an advertising company's reputation and revenue.

X. Xia et al. reported their work on developing a monitoring and prewarning system for accidents in the coal mines using data collected by a network of wireless sensors. They proposed a new data aggregation strategy and fuzzy comprehensive assessment model to derive useful information based on the collected data.

Acknowledgments

The guest editors would like to thank the authors for contributing to this special issue and thank all the reviewers for their time and rigorous reviews.

*Wenbing Zhao
Longxiang Gao
Anfeng Liu*