

EDITORIAL

Editorial for IEEE Access Special Section on Theoretical Foundations for Big Data Applications: Challenges and Opportunities

Big data is one of the hottest research topics in science and technology communities, and it possesses a great application potential in every sector for our society, such as climate, economy, health, social science, and so on. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, and manage. We can conclude that big data is still in its infancy stage, and we will face many unprecedented problems and challenges along the way of this unfolding chapter of human history.

It is critical to explore theoretical perspective of big data to efficiently and effectively guide its applications. We have witnessed the significant development in big data from various communities, such as the mining and learning algorithms from the artificial intelligence community, networking facilities from network community, and software platforms for software engineering community. However, big data applications introduce unprecedented challenges to us, and the existing theories and techniques have to be extended, upgraded to serve the forthcoming real big data applications, we even need to invent new tools for big data applications. For example, the appearance of big data forces us to study the events with a low or extremely low probability in statistics, which is usually a seldom studied area in history.

Therefore, we were motivated to organize the Special Section in IEEE ACCESS on Theoretical Foundations for big data Applications: Challenges and Opportunities. In response to the open call, we were pleased to see many submissions from different research communities all over the world. After a rigorous peer review process, we accepted thirteen papers, which investigate the theoretical aspects and application perspective of big data.

In the theoretical category, we have the following five papers.

With the advances of Artificial Intelligence (AI) techniques, research is flourishing in the area of autonomous systems. We therefore invited Abbass et al. to present a position paper (A review of theoretical and practical challenges of trusted autonomy in big data) discussing the challenges in establishing trust in autonomous systems. Using a big-date lens, the work discusses human-machine sensors, different computational models of trust, and architectures for trusted autonomous systems. It emphasises the need for models for

identity, intent, emotion, risk, and complexity management.

Many applications based on trajectories bring both unprecedented opportunities and great challenges. Zhenni Feng and Yanmin Zhu (A survey on trajectory data mining: Techniques and applications) explore various applications of trajectory data mining, e.g., path discovery and movement behavior analysis. They further review existing mining techniques and discuss them in a multi-layer framework which is a useful guideline for future solutions.

In the paper “Social set analysis: A set theoretical approach to big data analytics”, Vatraru et al. propose a new approach to big data analytics called social set analysis based on the sociology of associations and the mathematics of set theory. Implications for big data analytics, current limitations of the set theoretical approach, and future directions are also outlined.

Large-scale matrix inversion is a fundamental, but tough task for various emerging big data applications. In the work by Liu et al. (Spark-based large-scale matrix inversion for big data processing), a LU decomposition-based block-recursive algorithm is proposed. The evaluation results show that it can be a solid foundation to build a high-performance linear algebra library for big data processing

The paper by Willie K. Harrison (The role of graph theory in system of systems engineering) highlights some of the opportunities and challenges facing system of systems engineering that can be satisfied using graph-theoretical concepts and algorithms. Written in the style of a tutorial, the work summarizes existing approaches, and presents novel ideas for using graph theory to design, model, optimize, and deploy systems of systems in real time.

We have eight papers fall in to the application class as follows.

Zhaoyang Zhang, Hua Fang, and Honggang Wang (A new MI-based visualization aided validation index for mining big longitudinal Web trial data) propose a MI-based visualization aided validation index (MIVOOS) to determine the optimal number of clusters for big incomplete longitudinal web trial data with inflated zeros. Compared with its counterparts, the proposed MIVOOS shows its robustness in validating big web trial data under different missing data mechanisms using real and simulated web trial data.

Mehmood et al. (Protection of big data privacy) provide a comprehensive overview of the infrastructure of big data and the state-of-the-art privacy-preserving mechanisms in each stage of the big data life cycle. They further discuss the challenges and future research directions related to privacy preservation in big data.

Xiao et al. (A mobile offloading game against smart attacks) notice the vulnerability of mobile offloading, and propose an analytical model with three players to represent the problem. Nash and Stackelberg equilibria of the offloading game are derived and their existence conditions are discussed. Simulation results show that the proposed offloading strategy can improve the utility of the mobile device and reduce the attack rate of smart attackers.

In the paper "A tutorial on secure outsourcing of large-scale computations for big data" by Salinas et al., the authors review the recent advances in the secure outsourcing of large-scale computations for big data analysis. They focus on two most fundamental computational problems, linear algebra and optimization, and explain how to exploit recent developments in data privacy preserving techniques to construct secure outsourcing algorithms for large-scale computations.

Xu et al. (Exploiting trust and usage context for cross-domain recommendation) propose a novel method to solve the cross-domain recommendation problem. They first apply trust relations for predicting coarse ratings pertaining to cross-domain items. Then a new rating matrix is built from different domains. Finally, they compute the similarities of items and use item-based collaborative filtering to generate recommendations.

Energy is one of the most important parts in human life. To obtain a better understanding of the big data application

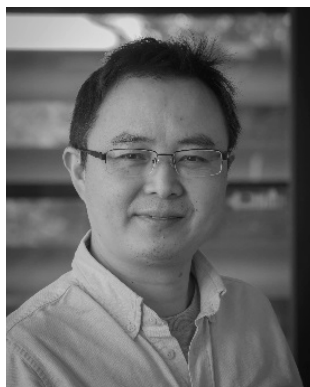
on energy, in the work by Jiang et al. (Energy big data: A survey) gives an overview of energy big data covering the recent researches and development in the context of an integrated architecture, key enabling technologies, security, typical applications and challenges.

He et al. (Big data analytics in mobile cellular networks) introduce a unified data model based on random matrix theory and machine learning. Then, big data analytics is introduced to apply in mobile cellular networks. Several illustrative examples are described, including analysing big signalling data, big traffic data, big heterogeneous data, and so on. Several open research challenges are also presented.

In the work by Chen et al. (A parallel patient treatment time prediction algorithm and its applications in hospital queuing-recommendation in a big data environment), a Patient Treatment Time Prediction (PTTP) model for each patient in the current task queue is predicted from large-scale and historical data. Then, a Hospital Queuing-Recommendation (HQR) system is developed based on the predicted time to recommend an effective treatment plan for patients to minimize their wait times in hospitals.

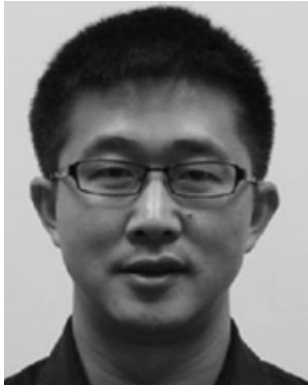
We would like to thank all the authors who submitted their research work to this Special Section. We would also like to acknowledge the contribution of many experts in this field who have participated in the review process, and offered comments and suggestions to the authors to improve their work. In particular, we would like to express our sincere appreciation to the Editor-in-Chief, Managing Editor, and the staff of IEEE ACCESS for their constructive suggestions, timely guidance, and professional support during the life cycle of this Special Section.

Finally, we hope our readers will enjoy the articles in this collection, and further explore in this promising and uncharted land.



SHUI YU (SM'12) is currently a Senior Lecturer with the School of Information Technology, Deakin University. He is a member of the Deakin University Academic Board from 2015 to 2016, AAAS, and ACM and the Vice Chair of the Technical Subcommittee on Big Data Processing, Analytics, and Networking of the IEEE Communication Society. His research interest includes networking theory, big data, and mathematical modeling. He has published two monographs, edited two books on big data and more than 150 technical papers, including top journals and top conferences, such as the IEEE TPDS, the IEEE TCC, the IEEE TCSS, the IEEE TC, the IEEE TIFS, the IEEE TMC, the IEEE TKDE, the IEEE TETC, and the IEEE INFOCOM. He initiated the research field of networking for big data in 2013. His h-index is 22. He actively serves his research communities in various roles. He served the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS as an AE from 2013 to 2015, and is currently serving the editorial boards of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS (Exemplary Editor for 2014), the IEEE ACCESS, the IEEE INTERNET OF THING JOURNAL, the IEEE COMMUNICATIONS LETTERS, and a

number of other international journals. Moreover, he has organized several Special Issues either on big data or cybersecurity. He has served on over 70 international conferences as a member of the organizing committee, such as the Publication Chair of the IEEE Globecom 2015 and the IEEE INFOCOM 2016, the TPC Co-Chair of the IEEE BigDataService 2015 and the IEEE ITNAC 2015, and as the Executive General Chair of the ACSW 2017.



CHONGGANG WANG received the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2002. He is currently a Member Technical Staff/Senior Manager of InterDigital Communications, Conshohocken, PA, USA, where he leads a research team on Internet of Things (IoT) Protocols and Big Data Innovation and Standardization. His current research interests include IoT, mobile communication and computing, and big data management and analytics. He is as a Distinguished Lecturer of the IEEE Communication Society from 2015 to 2016. He serves as the Founding Editor-in-Chief of the IEEE INTERNET OF THINGS JOURNAL and as a member of the editorial board for several journals, including the IEEE TRANSACTIONS ON BIG DATA and the IEEE ACCESS. He is also on the advisory board of IEEE-The Institute.



KE LIU received the B.S. degree from the Department of Automatic Control, Huazhong University of Science and Technology, and the Ph.D. degree in automatic control theory and its application from the Institute of Automation, Chinese Academy of Sciences. He is the Director of the Division of Computer Science, Directorate of Information Sciences, National Natural Science Foundation of China. He served as a Referee, an Editor, or a Guest Editor for several academic journals, and also served on the Technical Program Committee or the Organization Committee for a number of international conferences. He has successfully completed over a dozen of national projects as Leader or Core Member on discrete and hybrid system modeling and analysis and computer engineering. He has published more than 40 technical papers, as well as a number of articles on technical review and science administration.



ALBERT Y. ZOMAYA (S'88–M'91–SM'97–F'04) is currently the Chair Professor of High Performance Computing and Networking and an Australian Research Council Professorial Fellow with the School of Information Technologies, The University of Sydney. He is also the Director of the Centre for Distributed and High Performance Computing, which was established in 2009. He has authored or co-authored seven books and over 400 publications in technical journals and conferences, and has edited of 14 books and 17 conference volumes. He is currently the Editor in Chief of the IEEE TRANSACTION ON COMPUTERS and serves as an Associate Editor for 19 journals including some of the leading journals in the field. He was the Chair the IEEE Technical Committee on Parallel Processing from 1999 to 2003 and currently serves on its Executive Committee. He also serves on the advisory board of the IEEE Technical Committee on Scalable Computing and the advisory board of the Machine Intelligence Research Labs. He served as the General and Program Chair for over 60 events and served on the committees of more than 600 ACM and IEEE conferences. He delivered over 130 keynote addresses, invited

seminars, and media briefings. He is a fellow of the AAAS, the Institution of Engineering and Technology, U.K., a Distinguished Engineer of the ACM, and a Chartered Engineer. He received the 1997 Edgeworth David Medal from the Royal Society of New South Wales for outstanding contributions to Australian Science. He was also a recipient of the IEEE Computer Society's Meritorious Service Award and Golden Core Recognition in 2000 and 2006, respectively. Also, he received the IEEE TCPP Outstanding Service Award and the IEEE TCSC Medal for Excellence in Scalable Computing, both in 2011. His research interests are in the areas of parallel and distributed computing and complex systems.

...