Editorial

Junsheng Zhang*

Institute of Scientific and Technical Information of China,
Beijing, 100038, China
Email: zhangjs@istic.ac.cn
*Corresponding author

Yunchuan Sun

Business School,
Beijing Normal University,
Beijing, 100875, China
Email: yunch@bnu.edu.cn

Biographical notes: Junsheng Zhang is the Leader of the Data Mining and Information Analysis Research Group, at the Research Center for Information Science Theory and Methodology, Institute of Scientific and Technical Information of China (ISTIC). He received his PhD in Computer Science in 2009 from the Institute of Computing Technology, Chinese Academy of Sciences. His fields of interest include data and information analysis, information and knowledge management, semantic analysis and computing, mobile and cloud computing.

Yunchuan Sun is an Associate Professor in Beijing Normal University, Beijing, China. He acts as the Secretary of the IEEE Communications Society Technical Subcommittee for the Internet of Things from January 2013. He has been an Associate Editor of the *Springer Journal Personal* and *Ubiquitous Computing*. He received his PhD in Computer Science in 2009 from the Institute of Computing Technology, Chinese Academy of Science, Beijing, China. His research interests include internet/web of things, semantic link network, knowledge representation, cyber-physical-socio intelligence, and big data for the internet of things.

With recent advances in information and communication technologies, it is critical to improve the efficiency and accuracy of modern data processing techniques for information analysis and information service. During the past decade, we have witnessed tremendous technical advances in (wireless) sensor networks, the internet/web of things, cloud computing, mobile computing, and spatial/temporal data processing that lead to the advent of the big data era.

Big data is an emerging paradigm applied to datasets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. The datasets of big data are often from massive, various, heterogeneous and unstructured sources, such as social media, sensors, scientific applications, surveillance, video and image archives, web documents, medical records, business transactions and web logs; and the datasets are of large size with fast data in/out characteristics. Furthermore, big data has to be of high value and establish trust in it for business decision-making.

Big data has provided new challenges for data processing techniques. Various technologies are being discussed to support the handling of big data, such as massively parallel processing databases, scalable storage systems, cloud computing platforms, and MapReduce. Big data is more than simply a matter of size; and it is an opportunity to find insights in new and emerging types of data and content, to make business more agile, and to realise more intelligent applications. Therefore, massive data mining and analysis become the key technologies for using and understanding big data.

This special issue focuses on these topics, which form the essence of big data oriented sciences, technologies and applications, and is in collaboration with the international events: IIKI2015 (International Conference on Identification, Information and Knowledge in the Internet of Things 2015) in China (http://business.bnu.edu.cn/IIKI2015/) and COINFO 2015 (http://coinfo.istic.ac.cn). After two rounds of reviewing, the highest quality manuscripts were accepted for this special issue. At the end of the review process, we accepted 13 papers for this issue. Each of the papers was peer-reviewed by at least two experts in the field. In the following, we provide a brief introduction for each paper classified by research topic.

1 Network analysis algorithm and model

‘Time constraint influence maximization algorithm in the age of big data’ by Han et al. proposes a time constraint algorithm to maximise the influence in a large-scale network.

‘Analysing user retweeting behaviour on microblogs: prediction model and influencing features’ by Li et al.

analyses user retweeting behaviour on blogs from two aspects, including behaviour prediction and influencing feature selection.

2 Mining association rules from data

'Multi-criteria decisional approach for extracting relevant association rules' by Ait-Mlouk et al. studies the association rule extraction methods with multi-criteria for decision-making.

'Efficient attribute selection strategies for association rule mining in high dimensional data' by Harikumar et al. introduces the association rule mining in high dimensional data by adopting efficient attribute selection strategies.

3 Multimedia and image processing

'A robust video watermarking scheme using sparse principal component analysis and wavelet transform' by Shankar and Yamuna introduces a robust video watermarking scheme that uses sparse principal component analysis and wavelet transform methods.

'Optimisation for video watermarking using ABC algorithm' by Madhavan and Yamuna uses ABC algorithm to optimise video watermarking.

'DWT-based grey-scale image watermarking using area of best fit equation and cuckoo search algorithm' by Madhavan and Yamuna introduces grey-scale image watermarking using area of best fit equation and cuckoo search algorithm.

4 Natural language processing and text analysis

'Term extraction and correlation analysis based on massive scientific and technical literature' by Zeng et al. introduces term extraction and correlation analysis on massive scientific and technical literature.

'Towards patent text analysis based on semantic role labelling' by Li et al. uses the semantic role labelling technique to analyse massive patent text analysis.

5 Applications

'System architecture of coastal remote sensing data mining and services based on cloud computing' by Li et al. introduces a system architecture of coastal remote sensing data mining and services that is based on cloud computing.

'Collating multisource geospatial data for vegetation detection using Bayesian network – a case study of Yellow River Delta' by Gao et al. uses Bayesian network to collate multisource geospatial data for vegetation detection.

'A comparative study on disease risk model in exploratory spatial analysis' by Liu et al. introduces the comparative study on disease risk model for exploratory spatial analysis.

'Hybrid fuzzy collaborative filtering: an integration of item-based and user-based clustering techniques' by Yadav and Tyagi introduces a hybrid fuzzy collaborative filtering by combining the item-based and user-based clustering techniques.

Finally, we would like to take this opportunity to thank all authors for their valuable contributions to this issue. We are indebted to the anonymous reviewers for their hard work which has helped the authors to further enhance the quality of the manuscripts. We also would like to express our gratitude to Professor Kuan-Ching Li, the Editor-in-Chief of the journal, for giving us the opportunity and honour to serve as the guest editors of this issue.