

# Guest Editorial Preface

## Special Issue on Big Data and Services

Zibin Zheng, School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China

Mingdong Tang, School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan, China

Shangguang Wang, Beijing University of Posts and Telecommunications, Beijing, China

In recent years, big data has emerged as a widely recognized trend, attracting great attentions from government, industry and academia. Big data are high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization. However, due to the high volume, high velocity, and high variety (or other special) characteristics of big data, it poses enormous challenges to traditional information technologies from data storage, data transportation, data mining to data analysis. Big data also brings significant challenges to service-oriented computing, which is a promising information technology for distributed system development and integration. Actually, the past years have witnessed a significant convergence of big data and service-oriented computing. On one hand, the data amount generated by various services such as web services, cloud services, and social network services, is overwhelming and is still growing rapidly. On the other hand, it becomes more and more important to encapsulate every phase of the big data lifecycle as a service (e.g., data storage, data transportation, data mining, data analysis, etc.) and create values.

This special issue on big data and services focuses on addressing the challenging issues concerning the above two aspects. The special issue includes the following topics: Big Data Infrastructure-as-a-Service, Big Data Platform-as-a-Service, Big Data Analytics-as-a-Service, Business Aspects of Big Data-as-a-Service, Service Generated Big Data Mining, Analysis and Application, Big Data-Related Service Development, Deployment, Discovery and Selection. We launched an open call for submissions to this special issue In April, 2014 and received more than 20 submissions. The following five articles were selected through a rigorous review process:

- In the first paper, “A Method for Predicting Wikipedia Editors’ Editing Interest Based on a Factor Graph Model”, Haisu Zhang et al. propose an Interest Prediction Factor Graph (IPFG) model that integrates several important features related to interest prediction, including editors’ editing histories and social properties, hyperlinks, categories of entries, etc. They employ the gradient descent algorithm and the Loopy Sum-Product algorithm to estimate the parameters of the IPFG model, and the experimental result on a dataset collected from Wikipedia shows that the prediction accuracy is significantly higher than traditional collaborative filtering methods;
- In the second paper, “A New Symbolization and Distance Measure based Anomaly Mining Approach for Hydrological Time Series”, Pengcheng Zhang et al. study the problem of the abnormal hydrological time sequence mining. Combining with the field of hydrology, they explore the effective and accurate approach of abnormal pattern mining. Traditional abnormal

time pattern mining, especially in the field of hydrology, is based on distance measurement. This paper combines symbolization (FP\_SAX) of time series with distance measurement (SD\_DTW) to develop a new efficient approach suitable for feature extraction of hydrological time series and abnormal model identification;

- In the third paper, “An Efficient MapReduce Computing Model for Imprecise Applications”, Changjian Wang et al. propose an efficient computing model, named MapCheckReduce, to optimize the performance of MapReduce for imprecise applications. MapCheckReduce can terminate the map process of a job when some conditions are satisfied by reforming the map process in MapReduce. Besides the original data-processing on mappers, a new check process is added. The check process receives and analyzes messages from mappers and then determines whether to terminate the map process and start the data-processing in the reduce tasks;
- In the fourth paper, “Building and Analyzing of Enterprise Network: A Case Study on China Automobile Supply Network”, Liqiang Wang et al. propose a model of Enterprise Network (EN), which is a multi-level hypergraph with enterprises, employees, products and other related entities. They refine the EN model and present the foundation of EN to support social businesses and introduce a case study on the China Automobile Supply Network (CASN). The structure of the CASN is well analyzed and some special features such as power-law degree distribution and small-world are revealed. Some interesting patterns are also presented;
- In the fifth paper, “Moving Objects Gathering Patterns Retrieving based on Spatio-Temporal Graph”, Junming Zhang et al. propose a moving object gathering pattern retrieving method that aims to support the retrieving of gathering patterns based on spatio-temporal graph. In this method, they firstly use an improved R-tree based density clustering algorithm (RT-DBScan) to index the moving objects and collect clusters. Then, they maintain a spatio-temporal graph rather than storing the spatial coordinates to obtain the spatio-temporal changes in real time. Finally, a gathering retrieving algorithm is developed by searching the maximal complete graphs which meet the spatio-temporal constraints.

## ACKNOWLEDGMENT

The guest editors would like to thank all of the authors for submitting their latest research results to this special issue. They are especially grateful to the reviewers for their insightful and thoughtful reviews.

*Zibin Zheng is an associate professor in the School of Data and Computer Science at Sun Yat-sen University, China. He received ACM SIGSOFT Distinguished Paper Award at ICSE2010, and Best Student Paper Award at ICWS2010. His research interests include service computing and cloud computing.*

*Mingdong Tang is an associate professor in the School of Computer Science and Engineering at Hunan University of Science and Technology, Xiangtan, China. He received his PhD degree in Computer Science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2010. His current research interests include services computing, cloud computing, and data science. He is a member of IEEE and ACM.*

*Shangguang Wang received the PhD degree in computer science from the Beijing University of Posts and Telecommunications, Beijing, China, in 2011. He is an Associate Professor with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. His research interests include service computing, cloud services.*