Knowl Inf Syst (2014) 41:247–249
DOI 10.1007/s10115-014-0792-5

EDITORIAL

Special issue on big data research in China

Nanning Zheng · Jun Zhang · Chenghong Wang

On March 5–7, 2013, the National Natural Science Foundation of China (NSFC) organized the 89th Shuangqing Forum in Tongji University (Shanghai, China) on "Challenging Scientific Problems in Big Data Technologies and Applications" [1]. Big data refers to dynamic information that is generated in complex systems and has the characteristics of huge quantity, continuous sampling, multiple sources, and sparse values [2]. Big data has attracted tremendous attention from academia, industry, and the government both within China and internationally. Big data research seeks to extract useful information from massive data and to use it to facilitate our decision making [3]. In the future, big data technologies are expected to make a full use of public data resources to realize digital and intelligent transformations in areas such as traffic management, logistics, health care, and education [4]. However, the research on big data technologies still has many challenges. Therefore, this Springer KAIS special issue invites original research work on Big Data in China from both the invited speakers from the above NSFC forum and other established researchers in this field. We aim to bring together innovative designs, revolutionary ideas, and emerging applications of big data efforts.

For this special issue, from all invited and regular submissions, only nine were accepted for publication. The first paper, by Xu et al., proposed an automatic annotating technique based on the analysis of streaming social interactions of media content. This method first iteratively loads the streaming records to build the preference-sensitive subgraphs, then extracts static

N. Zheng (✉)
Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, 28 Xianning West Road, Xi'an 710049, Shaanxi, China
e-mail: nnzheng@mail.xjtu.edu.cn

J. Zhang
School of Advanced Computing, Sun Yat-sen University, Guangzhou 510275, China
e-mail: issai@mail.sysu.edu.cn

C. Wang
Division of Information Technology, National Science Foundation of China, 83 Shuangqing Road, Haiding District, Beijing 100085, China
e-mail: wangch@nsfc.gov.cn

and topologic features to describe these subgraphs, and finally, integrates these features into a learning-to-rank framework for automatic annotating. The experiments conducted by the authors show that the proposed approach had a highly competitive performance.

In the second paper, Li et al., studied the skyline query problem over uncertain data streams. The authors proposed a distributed parallel framework and a parallel streaming skyline approach to support efficient parallel execution of skyline computation. The effectiveness and efficiency of the proposed techniques were demonstrated by extensive experiments with real deployment over synthetic and real data.

The third paper, by Yuan et al., dealt with the query processing problems over huge RDF datasets. The study focuses on how to execute RDF queries efficiently. The authors presented a query processing approach highlighted by dynamic plan generation and pipelining execution. Additionally, optimization techniques are employed to further improve the performance of the query processing. This approach is compared to existing query engines on three RDF datasets of over a billion triples. In all cases, the proposed approach exhibits a competitive performance.

Chen et al., looked, in the fourth paper, into ways to support in-memory big data analytics. The authors proposed a hybrid memory which combines DRAM and flash-based solid-state drive (SSD) together. The main idea is to take SSDs as the assistant of DRAM to configure a large as well as economical main memory. Experimental results demonstrate that the hybrid memory is able to extend the DRAM by more than twice.

Lu et al., presented in the fifth paper a quality prediction method for cold rolling process with varying operating condition by considering the 3Vs (volume, velocity, and variety) of the cold rolling process data. In this method, locality preserving projection (LPP) is used for dimensionality reduction, and mixture probabilistic linear regression (MPLR) is used to develop a global prediction model. The authors conducted extensive experiments to illustrate the advantages of their proposed approach.

Zhang et al., presented in the sixth paper a computational skew-tolerant approach that aims to tolerate the negative effects of unresolved computational skews. This approach divides a large computational task into a number of sub-tasks, which depends on the states of only a few objects from the previous iteration. The experiments conducted by the authors show that this approach had a highly competitive performance in comparison with the state-of-the-art solutions.

The seventh paper, by Liu et al., proposed a novel data-based adaptive online prediction model for plant-wide production indices. This model combines classical online support vector regression (SVR) and Cherkassky's method. Moreover, the authors use dynamically determined penalty factors to improve the prediction precision. Based on the experimental results, the authors concluded that their proposed approach was effective and performed better than existing methods.

In the eighth paper, Liu et al., dealt with the problem of data security in scientific cloud workflows. The authors proposed a security-aware intermediate data placement strategy for scientific workflows in cloud. This strategy uses an ACO-based algorithm to dynamically place the intermediate data to improve data security. Experimental results show that the proposed strategy can improve the intermediate data security while the data transfer time is guaranteed.

Finally, in the ninth paper, He et al., studied the issue of data compression in wireless sensor networks (WSNs) environment. The authors proposed an algorithm for the compression of spatial–temporal data and recovery of sensor data in a WSN deployed in an underground tunnel. Experimental studies using data acquired from Shanghai Metro Line 12 demonstrated the effectiveness of the proposed algorithm.

The nine papers included in this special issue are representative of the current trends in the big data research field in China. We hope that these efforts can provide insights into other researchers for their studies in the years to come.

References

1. Wang C, Chen W, Zhang J, Song S, Lu R (2014) Challenging scientific problems in big data technologies and applications. Bull Natl Nat Sci Found China 2014(2):92–98
2. Mayer-Schönberger V, Cukier K (2013) Big data: a revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt, Boston, MA
3. National Research Council (2013) Frontiers in massive data analysis. The National Academies Press, Washington D.C.
4. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, Byers AH (2011) Big data: the next frontier for innovation, competition, and productivity. The McKinsey Global Institute, USA