Preface of Special Issue on BigDataSE 2016

Heng Qi[1,2] and Keqiu Li[1]

[1] School of Computer Science and Technology, Dalian University of Technology, China
[2] Graduate School of Information Science, Nagoya University, Japan. Email: hengqi@dlut.edu.cn, keqiu@dlut.edu.cn

1. INTRODUCTION

With the development of information technology, the explosion of data is produced in our lives. It is a big challenge to capture, manage, and process more and more data within a tolerable elapsed time. Moreover, it is very difficult to acquire more useful information from very large data sets. Against this background, big data is proposed as an emerging paradigm drawing more and more attentions from the academia and industry in recent years. BigDataSE is the acronym of the International Conference on Big Data Science and Engineering. This conference is created to provide a prime international forum for researchers, industry practitioners and environment experts to exchange the latest fundamental advances in the state of the art and practice of big data and broadly related areas.

This special issue presents the recent advances in big data and related areas, which are selected out from papers presented at the BigDataSE 2016 [1]. The process of selection is very rigorous. Nine papers are selected based on their originality, technical quality and significance. The topics of these papers include several typical applications in big data, such as social network, recommendation system, image processing, information retrieval, mobile application and Cyber Physical Systems (CPS). The main contributions of these papers are as follows:

- Tianqing Zhu et al proposed a method of graph update in online social networks to achieve differentially privacy when releasing a large set of queries [2].

- Kehong Zhang et al proposed one algorithm named TCRQDG to address the reachability querying problem for large scale graph in the social network [3].

- Chen Yu et al presented an efficient trajectory data processing model to mine hot route from a lot of trajectory logs [4].

- Chunjing Xiao et al proposed a novel tourist routes recommendation method, which can adapt to the characteristics of collected data effectively [5].

- Lufeng Yuan et al studied the certain events detection in biomedical imaging. By their proposed method, the flash event in fluorescence images can be detected efficiently [6].

- Wenyu Qu et al designed an effective and efficient algorithm to produce the skeletons of sparse shapes in a large number of images. This algorithm can be applied into pattern recognition for big image data [7].

- Yong Zhang et al presented a new parallel XML keyword search algorithm based on MapReduce programming model. The proposed algorithm adapts to keyword search of big XML data achieving high execution efficiency [8].

- Lin Wang et al propose a novel design that using the multi-layer filtering of channel state information (CSI) to identify moving targets in dynamic environments and analyze the gait periodicity of human [9].

- Junfeng Xu et al designed and implemented online dynamic protection framework for Android application, which can protect Android application from massive malicious attacks [10].

- Tingting Yang et al developed effective algorithms to achieve optimal resource scheduling for big video data transmission in CPS, by which the video packets can be successfully transmitted before their deadlines [11].

2. THEMES OF THE SPECIAL ISSUE

2.1 Social Network and Recommendation System

With the popular of social network application, the volumes of data collected in these applications have become larger and larger. Generally, the data collected from the social network is in the form of graphs. So some problems of social network can be abstracted to mathematical models in graph theory. The paper entitled "An Iteration-based Differentially Private Social Network Data Release" focuses on the privacy models of graph data [2]. In this paper, the authors transfer the query release problem to an iteration based training process, thereby generating a synthetic graph to answer queries accurately. Compared to existing work, the proposed iteration based method can achieve differential privacy of nodes and edges when releasing a large set of queries for graph data. The paper entitled "The Optimization Reachability Query of Largescale Multiattribute Constraints Directed Graph" focuses on the problem of reachability on large scale graphs [3]. In social network applications, the authors find a issue of the multi-attribute constraints reachability on directed graph. To address this issue, they propose an algorithm named TCRQDG, which can achieve fast and accurate query on large scale graph data.

In the big data era, recommendation system is a very hot topic. Accord to big data analysis, we can predict the user preferences thereby recommending some useful information to users. Two papers in this special issue are about hot route recommendation and tourist routes recommendation, respectively. The location-based services are emerging technology with the rapid expansion of mobile devices. The paper entitled "Tensor-based User Trajectory Mining" aims to realize the hot route mining based on the individual's trajectory logs on the mobile devices [4]. In this paper, the authors present a collaborative tensor calculation method to process trajectory data. By the trajectory data processing model, they can compute the recommended hot route with considerable accuracy. The paper entitled "Research on Tourist Routes Recommendation based on the User Preference Drifting Over Time" aims to realize the tourist routes recommendation based on the tourism data from airlines or other tourism companies [5]. The authors propose a route recommendation model based on dynamic dividing a user's interaction history, which is more adapt to the high sparsity and temporal characteristics of tourist data.

2.2 Image Processing and Information Retrieval

Recently, multimedia applications become more and more popular. Images are increasingly being considered as typical examples of big data. Researchers have to face new challenges in image processing. The paper entitled "Automated and Precise Event Detection Method for Big Data in Biomedical Imaging with Support Vector Machine" focuses on biomedical image processing in medical area [6]. The authors analyze and implement an improved SVM based method to detect and analyze important biological events in the Escherichia coli images. The ac-

curacy of the proposed method achieves an F-value of about 0.81. The paper entitled "Effective Piecewise Linear Skeletonization of Sparse Shapes" focuses on the skeletonization algorithms of shape analysis and recognition [7]. As we know shape recognition is one of the fundamental problem of pattern recognition and image processing. Existing skeletonization algorithms fail to extract correct skeleton, because they utilize the connectivity of image pixels inside the object regions. Different from these existing methods, the authors propose a new skeletonization algorithm based on node sampling and edge contraction, which easily captures closed topological property and avoids error links between different objects.

Big data also poses immense challenges to information retrieval. It is difficult to search massive datasets by keywords in realtime, especially the data with complex structure. Nowadays, XML data is commonly used in Internet. It is difficult to search massive XML dataset by keywords in real time. In the paper entitled "MapReduce Implementation of an Improved XML Keyword Search Algorithm", the authors propose a kind of parallel XML keyword search algorithm, in which the smallest lowest common ancestor (SLCA) is calculated based on intelligent grouping [8]. They also implement the proposed algorithm based on MapReduce programming model to show that the proposed algorithm can process SLCA-based keyword search of large-scale XML data. With the development of Internet of Things (IoT) and wireless communication techniques, more and more devices are connected to the Internet via WiFi, 5G, etc. The massive CSI data gathered at the receiver contains a lot of valuable information that can be used for various purpose, such as human movement detection, monitoring, indoor/outdoor location. The paper entitled "Human Movement Detection and Gait Periodicity Analysis Via Channel State Information" is dedicated to analyze the gait periodicity of human walking [9]. The authors leverages the principal component analysis (PCA) and discrete wavelet transform (DWT) multi-layer filtering algorithm for detecting features of CSI subcarriers, the feature difference detected from CSI subcarriers are then used to analyze the gait periodicity of human walking in the CSI streams. The experimental results indicate higher robustness can be achieved in different scenes with their proposed methods.

2.3 Mobile Application and Cyber Physical Systems (CPS)

Nowadays, mobile devices are necessary for our daily lives. However, the security of mobile devices need to be improved to deal with the malicious attacks in big data era. The paper entitled "A Dynamic Online Protection Framework for Android Applications" focuses on improving the security of Android system [10]. To achieve this goal, the authors propose a new online dynamic protection framework, in which the online keys can be exchanged to original APK file. Compared to the existing static protection framework, the proposed dynamic framework can prevent the android applications from cracking and malicious analysis to the utmost.

In recent years, Cyber Physical Systems (CPS) is emerging domain attracting increasing interest. However, CPS has to face the challenges from big data. For example, due to the limitation

of wireless transmission, it is difficult to deliver a large volume of monitoring videos in CPS. The paper entitled "Online and Offline Scheduling Schemes to Maximize the Weighted Delivered Video Packets Towards Maritime CPSs" aims to address the video data delivery scheduling problem in maritime CPS [11]. The authors formulate the resource allocation and scheduling problem by jointly consider the intermittent network connectivity and cooperative transmission scheme. Then, they propose online and offline algorithms to maximize the weights of uploaded video packets. From the evaluation results, we can see their method can obtain better performance.

Acknowledgment

REFERENCES

1. The 10th International Conference on Big Data Science and Engineering (BigDataSE-16), Tianjin, China. http://adnet.tju.edu.cn/BigDataSE2016/index.html.

2. Tianqing Zhu, Mengmeng Yang, Ping Xiong, Yang Xiang, and Wanlei Zhou. *An Iteration-based Differentially Private Social Network Data Release*. Computer Systems Science & Engineering, 2016.

3. Kehong Zhang and Keqiu Li. *The Optimization Reachability Query of Largescale Multiattribute Constraints Directed Graph*. Computer Systems Science & Engineering, 2016.

4. Chen Yu, Qinmin Hong, Dezhong Yao and Hai Jin. *Tensor-based User Trajectory Mining*. Computer Systems Science & Engineering, 2016.

5. Chunjing Xiao, Kewen Xia, Yongwei Qiao and Yuxiang Zhang. *Research on Tourist Routes Recommendation based on the User Preference Drifting Over Time*. Computer Systems Science & Engineering, 2016.

6. Lufeng Yuan, Erlin Yao and Guangming Tan. *Automated and Precise Event Detection Method for Big Data in Biomedical Imaging with Support Vector Machine*. Computer Systems Science & Engineering, 2016.

7. Wenyu Qu, Zhiyang Li, Junfeng Wu, Yinan Wu and Zhaobin Liu. *Effective Piecewise Linear Skeletonization of Sparse Shapes*. Computer Systems Science & Engineering, 2016.

8. Yong Zhang, Jing Cai and Quanlin Li. *MapReduce Implementation of an Improved XML Keyword Search Algorithm*. Computer Systems Science & Engineering, 2016.

9. Lin Wang, Zijuan Liu, Wenyuan Liu and Binbin Li. *Human Movement Detection and Gait Periodicity Analysis Via Channel State Information*. Computer Systems Science & Engineering, 2016.

10. Junfeng Xu, Zhang Li and Mao Ye. *A Dynamic Online Protection Framework for Android Applications*. Computer Systems Science & Engineering, 2016.

11. Tingting Yang, Hailong Feng, Chengming Yang, Ge Guo and Tieshan Li. *Online and Offline Scheduling Schemes to Maximize the Weighted DeliveredVideo Packets Towards Maritime CPSs*. Computer Systems Science & Engineering, 2016.