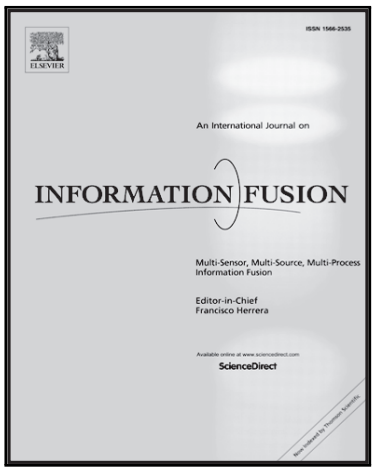

Big Data Fusion in Internet of Things

Zheng Yan , Jun Liu , Laurence T. Yang , Nitesh Chawla

PII: S1566-2535(17)30253-1
DOI: 10.1016/j.inffus.2017.04.005
Reference: INFFUS 867

To appear in: *Information Fusion*

Received date: 17 April 2017
Accepted date: 22 April 2017

Please cite this article as: Zheng Yan , Jun Liu , Laurence T. Yang , Nitesh Chawla , Big Data Fusion in Internet of Things, *Information Fusion* (2017), doi: 10.1016/j.inffus.2017.04.005

Big Data Fusion in Internet of Things

The Internet of Things (IoT) brings the real physical world, virtual cyber world and digital world together. Varieties of sensors, such as mobile termainals, cameras, microchips, wearables and even the Internet and socialized human beings, play an important role in IoT. These sensors collect, generate, and preserve a diversity of data with different representations, scales, and densities from various "things", which offers IoT the ability to measure, infer and understant environments. Integrating things, data and semantic opens opportunities for knowledge discovery, and further makes it possible to provide advanced and intelligent services.

Data (information) fusion is an essential and integral part of IoT. Data in IoT characterised by dynamic and heterogeneous leads to inadequacy of simple single-source analysis methods. Data fusion integrates multiple data and knowledge into a consistent, accurate and useful representation, in which the data are fused to high-quality information to provide a reliable decision support. Therefore, it is important to investigate techniques for understanding and resolving issues about data fusion in IoT.

However, there are significant barriers to overcome before the potential benefits are fully realized. First, data in IoT comes in large amounts, is a mixture of structured and unstructured information, arrives at speed and can be of uncertain provenance. Many existing solutions become improper due to high computation complexity. Second, data sources in IoT are often of different quality, and with significant differences in coverage, accuracy and timeliness of data, which brings significant challenges to achieve trustworthy data fusion and analytics. Third, managing, extracting and deeply understanding valuable knowledge from multi-modal sources in IoT is the biggest challenge. Big Data fusion in IoT (BDFIoT) calls for advanced techniques that can fuse the knowledge from various data sources organically and efficiently in a machine learning and data mining task.

This special issue aims at presenting advanced research results related to big data fusion in Internet of Things. We finally selected 7 papers from a total of 41 submissions after a rigorous review process and panel discussion on their novelty, research significance, technical correctness, evaluation comprehension and presentation.

The first paper, titled *A Delay-Aware Schedule Method for Distributed Information Fusion with Elastic and Inelastic Traffic* by *Shen et al.*, proposes an online scheduling algorithm and its distributed implementation, named Delay-Guaranteed CSMA, in order to guarantee the performance of Distributed Information Fusion (DIF). Both the timing constraints and the historical transmission statistics of sensors are taken into consideration to ensure good delay-guaranteed satisfaction and real-time data delivery.

The second paper, titled *CSF: Crowdsourcing Semantic Fusion for Heterogeneous Media Big Data* by *Guo* et al., proposes a novel solution named Crowdsourcing Semantic Fusion (CSF) that makes use of the collective wisdom of social users and introduces crowdsourcing computing into semantic fusion for overcoming the challenges that manual annotation is inefficient and automatic annotation is inaccurate. This work provides a convenient interface for users to extract semantic information given heterogeneous media documents obtained from the Internet and designs the algorithms to perform the normalization and fusion given heterogeneous semantic objects.

The third paper, titled *Using Check-in Features to partition Locations for Individual Users in Location Based Social Network*, by *Yu* et al., aims to categorize a location for a user once he or she makes initial check-in there. This research utilizes classification in machine learning to partition locations for individual users based on a publicly available check-in data set. An interesting finding of this study is the contribution of different feature categories varies in

location classification, where social features appears to offer the least contribution.

The fourth paper, titled *MRI Segmentation Fusion for Brain Tumor Detection*, by *Cabria and Gondra*, presents a new algorithm called Potential Field segmentation (PFS) and proposes using ensemble approaches to combine the results generated by PFS and other methods to achieve a fused segmentation. The purpose of this study is to improve the efficiency and accuracy of generating precise segmentations of brain tumors from magnetic resonance images (MRI).

The fifth paper, titled *Air Quality Data Clustering using EPLS Method*, by *Chen* et al., proposes a model-based feature extraction method to solve the problem of PM2.5 data clustering in order to overcome the impact of other factors in the air and relative high noise level and outlier. Through mode decomposition, dimension reduction and least squares projection, the proposed method can deal with high noise level and outlier air quality clustering problems.

The sixth paper, titled *High-order Possibilistic c-Means Algorithms Based on Tensor Decompositions for Big Data in IoT*, by Zhang et al., proposes two high-order possibilistic c-means algorithms based on the canonical polyadic decomposition (CP-HOPCM) and the tensor-train network (TT-HOPCM) for clustering big data in order to support low-end devices with limited memory space and computing power in IoT systems. The experimental results show their potential for big data clustering due to high compression rate, low storage cost and satisfactory clustering accuracy.

The seventh paper, *BFSI-B: An Improved K-hop Graph Reachability Queries for Cyber-Physical Systems*, by *Xie* et al. proposes a compound-index method, namely BFSI-B, which uses special index to decrease the k-hop reachability query time in a graph and improve pruning efficiency in order to achieve high scalability for being applied into very large graphs.

Editing this special issue took us great efforts due to the big amount of submissions, but it was an advanced and valuable experience. We would like to thank all authors and reviewers for their contributions and the support of professor Francisco Herrera, the editor-in-chief of Information Fusion. IoT big data fusion is a wide research field. We believe there are many other interesting issues that should be seriously explored and investigated, but unfortunately not covered in this special issue. We hope this special issue can greatly motive future research to achieve advanced research results in this field.

Acknowledgment

This work is sponsored by the National Key Research and Development Program of China (grant 2016YFB0800700), the NSFC (grants 61672410 and U1536202), the Project Supported by Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2016ZDJC-06), the 111 project (grants B16037 and B08038), the PhD grant of the Chinese Educational Ministry (grant JY0300130104), and Aalto University.

Guest Editors
Zheng Yan
State Key Laboratory of Integrated Services Networks, School of Cyber Engineering, Xidian University,
No. 2 South Taibai Road, 710071 Xi'an, China
Department of Communications and Networking, Aalto University,
Otakaari 5, 02150 Espoo, Finland
Tel.: +86 18691958048; fax: +358 50 4836664.

E-mail address: zyan@xidian.edu.cn, zheng.yan@aalto.fi

Jun Liu
Xi'an Jiaotong University, China
E-mail address: liukeen@mail.xjtu.edu.cn

Laurence T. Yang
St. Francis Xavier University, Canada
E-mail address: ltyang@gmail.com

Nitesh Chawla
University of Notre Dame, USA
Email: nchawla@nd.edu