



# Wireless Personal Communications: Machine Learning for Big Data Processing in Mobile Internet

Jun Guo<sup>1</sup> · Zheng-Hua Tan<sup>2</sup> · Sung Ho Cho<sup>3</sup> · Guoqiang Zhang<sup>4</sup>

Published online: 30 August 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

With the rapid development of the Internet, academic researches and industrial applications are ever progressing in numbers and generating a massive amount of multi-modal data from multiple sources. The most prominent example is the mobile Internet, where increasingly more data sets and data streams, such as image, audio, social media, are collected, transmitted, and analyzed. These big data collected from a mobile environment allow researchers to learn and understand significant characteristics that traditional data cannot provide. Hence, mobile big data bring us huge opportunities to understand our world more precisely as well as challenges to analyze, process, store, and transmit such a massive amount of data. In order to analyze and understand mobile big data, the developments of related theories, frameworks, models, and algorithms are crucial. The scope of this special issue is to provide machine learning-based theoretical foundations as well as novel models and algorithms to solve these challenges.

Welcome to this special issue of Machine Learning for Big Data Processing in Mobile Internet. The special issue focuses on presenting the latest research results in the field of machine learning for data processing in the mobile Internet, and disseminating contributions by world-wide leading researchers from both academia and industry addressing aspects ranging from case studies of particular problems to novel learning theories and approaches, including (but not limited to):

- Machine learning theory and methodology for mobile big data processing.
- Pattern and feature learning in mobile big data.
- Intelligent resource scheduling in heterogeneous mobile cloud computing.

---

✉ Jun Guo  
guojun@bupt.edu.cn

<sup>1</sup> Pattern Recognition and Intelligent System Lab., Beijing University of Posts and Telecommunications, No. 10 Xitucheng Rd., Haidian District, Beijing 100876, China

<sup>2</sup> Department of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7, 9220 Aalborg, Denmark

<sup>3</sup> Department of Electronic Engineering, Hanyang University, 222 Wangsimni-Ro, Seongdong-Gu, Seoul 04763, South Korea

<sup>4</sup> School of Computing and Communications, University of Technology Sydney, Ultimo, NSW 2007, Australia

- Machine learning-driven visualization for mobile big data.
- Infrastructure and platforms for mobile bit data processing.

The special issue on Machine Learning for Big Data Processing in Mobile Internet is composed of fourteen high-quality original papers covering theoretical and practical aspects on this emerging topic in order to provide novel ideas and directions for the relevant scientific industrial and academic community. All submitted papers have been reviewed by three independent reviewers.

This special issue opens with a paper entitled “Machine Learning Based Big Data Processing Framework for Cancer Diagnosis Using Hidden Markov Model and GM Clustering” by Gunasekaran Manogaran et al., where a Bayesian hidden Markov model (HMM) with a Gaussian mixture (GM) clustering approach is proposed to model the DNA copy-number change across the genome. The proposed Bayesian HMM with a GM clustering approach is compared to various existing approaches, such as the pruned exact linear time method, the binary segmentation method and the segment neighborhood method.

The second paper, entitled “Framework for Fast and Efficient Cloud Video Transcoding System using Intelligent Splitter and Hadoop MapReduce” by Duraipandy Kesavaraja and A. Shenbagavalli, proposes an intelligent video splitter, which uses the map reduce algorithm to provide efficiency based on the time factor. The important performance metrics, including video distortion (VD) and video distortion due to frame dependency (FDD), were considered. The results showed that the proposed framework perceptibly outperforms the prevailing strategies. Furthermore, this method may be extended to supply associate automatic device aware video standards.

The third paper, entitled “Cluster-PSO based Resource Orchestration for Multi-task Applications in Vehicular Cloud” by Qi Qi et al., formulates the VCC resource orchestration as an optimization problem and proposes a cluster-particle swarm optimization (PSO) algorithm to obtain the resource orchestration policy. The experiment results show that the cluster-PSO algorithm, compared to other PSO algorithms, is able to achieve a higher resource orchestration accuracy in an acceptable time. The performance of the cluster-PSO based resource orchestration is especially outstanding when there are more tasks in a given application and the given vehicle has more optional VCC resources.

The fourth paper, entitled “Mining Association Rules Based on Deep Pruning Strategies” by Lei Li et al., presents an algorithm called MAR-DPS, which has deep pruning strategies containing three methods to compress the size of frequent itemsets and reduce the joining numbers in generating new frequent itemsets. The proposed method is also able to select the appropriate method for generating frequent 2-itemsets when testing on different data sets. Extensive experimental results on three different data sets have demonstrated that this MAR-DPS algorithm performs much better than other tested algorithms.

The fifth paper, entitled “Positive and Negative Link Prediction Algorithm based on Sentiment Analysis in Large Social Networks” by Debasis Das, applies sentiment analysis in social networks to detect all relationships between the nodes, which categorizes users in five simple categories: highly positive, positive, neutral, negative and highly negative.

The sixth paper, entitled “External Expansion Risk Management: Enhancing Microblogging Filtering Using Implicit Query” by Zhen Yang et al., suggests the external document can be viewed as a complete statement of an explicit query, and encodes the filtering preferences with a diverge degree between the external document and the original explicit query. Thus the optimal filtering action is the one that allows one to trade off diverge degree against generalization performance. With respect to the established baselines, the proposed algorithm yields compelling results for meaningful tweet retrievals. In

addition, this work furthers the understanding of the innate risk characteristics of external expansion for the design of microblogging filtering systems.

The seventh paper, entitled “Machine Learning Based Resource Utilization and Pre-estimation for Network on Chip (NoC) Communication” by Adesh Kumar et al., presents the use of machine learning techniques to predict the FPGA resource utilization for NoC. The present study aids NoC chip planning prior to the designing of the chip itself by taking into account known hardware design parameters, memory utilization and timing parameters such as minimum and maximum period, frequency support etc. The developed model has been validated and has performed well on independent test data.

The eighth paper, entitled “A Human-in-the-loop Architecture for Mobile Network: from the View of Large Scale Mobile Data Traffic” by Yuanyuan Qiao et al., presents a Human-In-The-Loop architecture for mobile networks that discovers user needs of a network resource by understanding the data traffic usage behavior of users. Based on real data traffic of a mobile network, the authors analyze data traffic patterns of heavy and normal users from online browsing behavior in urban functional areas to explain how and why the data traffic is consumed. Then they propose a Latent Dirichlet Allocation model-based solution to correlate data traffic, user behavior, and urban ecology to gain deep insights into the spatio-temporal dynamic of data traffic usage behavior for different groups of users. The results of this work can potentially be used to help allocate network resources, improve the Quality of Experience according to the users’ needs, and even design the network of the future.

The ninth paper, entitled “Defect prediction in Android binary executables using deep neural network” by Feng Dong et al., proposes an approach called smali2vec, which is able to extract both token and semantic features of the defective files in apks. The new approach generates features that capture the characteristics of smali (decompiled files of apks) files in apks. Based on these features, deep neural network (DNN) are trained to form a defect prediction model in order to achieve a high accuracy. The model has been used in practical systems and helped locate many defective files in apks.

The tenth paper, entitled “Distributed and Adaptive Analog Coding for Video Broadcast in Wireless Cooperative System” by Yumei Wang et al., introduces a distributed and adaptive analog coding scheme called ACVC (adaptive cooperative video coding) based on AJSCC and with the concept of coset coding in distributed source coding, to improve the overall video broadcast quality in wireless cooperative systems. Particularly, an adaptive packet discarding module is introduced to the framework to avoid video quality deterioration under severe channel conditions. Furthermore, a model for quantization step selection of coset coding is built to minimize the redundancy in the cooperative signal and improve the noise-robustness ability of the video coding scheme. The experimental results show that, ACVC has stronger adaptability and thus obtains higher quality of broadcasted video than existing wireless cooperative schemes in the literature under different channel conditions.

The eleventh paper, entitled “Variational Bayesian Inference for Infinite Dirichlet Mixture Towards Accurate Data Categorization” by Yuping Lai et al., proposes a variational Bayesian learning approach to an infinite Dirichlet mixture model (VarInDMM), which inherits the confirmed effectiveness of modeling proportional data from an infinite Dirichlet mixture model. Based on the Dirichlet process mixture model, VarInDMM has an interpretation as a mixture model with a countably infinite number of components, and it is able to determine the optimal value of this number according to the observed data. By introducing an extended variational inference framework, the paper further obtains an analytically tractable solution to estimate the posterior distributions of the parameters for

the mixture model. Experimental results on both synthetic and real data demonstrate its good performance on object categorization and text categorization.

The twelfth paper, entitled “Entropy with Local Binary Patterns for Efficient Iris Liveness Detection” by Waleed S.-A. Fathy and Hanaa S. Ali, introduces an efficient system for detecting iris attacks. The system avoids the segmentation and the normalization stages employed traditionally in fake detection systems. Results show that more discriminative features can be obtained using the proposed system. System complexity and processing time are reduced noticeably, and the system is robust to different types of fakes.

The thirteenth paper, entitled “Optimization of Spatially-Coupled Multiuser Data Transmission through Machine Learning Methods” by Zhongwei Si et al., proposes a spatial coupling multiple access strategy to obtain higher spectral efficiency for multi-user data transmission. The optimization of the system can be carried out through, for example, constellation rotation or power allocation, which mitigates the interferences and distinguishes users/data streams better. The optimization aims for the maximization of the average mutual information between the transmitted data symbols and the received signal. To solve the optimization problem, the authors employ machine learning based methods to locate the possibly global optimal solution. Simulation results show that the optimization clearly contributes to the performance improvement of the system.

Finally, the fourteenth paper, entitled “A Survey on Long Term Evolution scheduling in data mining” by Divya Mohan and Geetha Mary Amalanathan, conducted intensive survey on long term evolution (LTE) scheduling in data mining. Scheduling procedures implemented in wireless networks consist of varied workflows, such as resource allocation, channel gain improvement, and reduction in packet arrival delay. Among these techniques, LTE scheduling is most preferable due to its high-speed communication and low-bandwidth consumption. LTE allocates resources to the workflow based on time and frequency domains. Normally, the information gathered prior to scheduling increases the processing time since each attribute of the users has to be verified. In order to solve this issue, parallel processing via data mining is analyzed in recent studies. The label that is assigned to the user attributes contributes primarily on scheduling time slots effectively. The label assignment and the parallel processing via data mining reduces the delay and increases the throughput, respectively. Additionally, the matched data extraction from the library and the prediction of available channels with fewer dimensions posed major challenges in the LTE scheduling. This paper surveys about various LTE scheduling algorithms, dimensionality reduction techniques, optimal feature selection techniques, multi-level classification techniques, and data mining combined with LTE techniques. A brief survey illustrates the impact of each technique on 3G/4G networks, channel availability prediction, scheduling of time slots in detail. A brief comparison of the techniques involved in the respective LTE processes via tabular form reveals that the verification of channel and user availability are the primary functions of the LTE scheduling. The survey of this paper identifies the limitations, such as computational complexity and poor scheduling performance in the existing systems and encourages researchers to develop novel algorithms for LTE scheduling.

In the end, we would like to thank all the authors for their contributions to this special issue. We also thank the reviewers for their constructive comments that help in improving the quality of the papers. We are very grateful to WPC Editorial Team for providing the opportunity to organize this special issue. We hope the related scientific community will benefit from this issue.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Jun Guo** received B.E. and M.E. degrees from Beijing University of Posts and Telecommunications (BUPT), China in 1982 and 1985, respectively, Ph.D. degree from the Tohoku-Gakuin University, Japan in 1993. At present he is a professor and a vice president of BUPT. His research interests include pattern recognition theory and application, information retrieval, content based information security, and bioinformatics. He has published over 200 papers on the journals and conferences including SCIENCE, Nature Scientific Reports, IEEE Trans. on PAMI, Pattern Recognition, AAAI, CVPR, ICCV, SIGIR, etc.



**Zheng-Hua Tan** is Professor and a co-head of the Centre for Acoustic Signal Processing Research (CASPR) at Aalborg University, Denmark. He was a Visiting Scientist at MIT, USA, an Associate Professor at SJTU, China, and a postdoctoral fellow at KAIST, Korea. His current research interests include machine learning, deep learning, pattern recognition, speech and speaker recognition, noise-robust speech processing, multimodal signal processing, and social robotics. He has served as an Associate/Guest Editor for several journals.



**Sung Ho Cho** received the Ph.D. degree in Electrical and Computer Engineering from the University of Utah, Salt Lake City, USA, in 1989. After finishing the graduate studies, he returned to South Korea, and joined Electronics and Telecommunications Research Institute (ETRI) of Korea for three years. He then joined the Department of Electronic Engineering in Hanyang University, Seoul, South Korea, in 1992, where he is currently a full professor. He served as Associate Dean and later Dean of the College of Engineering at Hanyang University from 2013 to 2017. Professor Cho has been actively engaged in numerous government and industry research projects related to applied signal processing, context-aware computing, and radar applications. Especially in 2017, based on his over-7-year research experiences on IR-UWB radar computing, he co-founded the two university start-up companies, Xandar Tech (<http://xandartech.com>) for security/safety solutions and Kardian (<http://kardian.com>) for healthcare solutions. He is currently working as a CTO at both companies. From 2008 to 2012, Professor Cho was appointed as a Regular High-Level Visiting Scientist at Beijing University of Posts and Telecommunications (BUPT) of China under the 111 Fellowship. Since 2013, he has been holding an invited researcher position at BUPT under the Recruitment Program of High-Level Foreign Experts supported by the Ministry of Education in China.



**Guoqiang Zhang** received the B. Eng. from University of Science and Technology of China (USTC) in 2003, M.Phil. degree from University of Hong Kong in 2006, and Ph.D. degree from Royal Institute of Technology in 2010. From the spring of 2011, he worked as a Post-doctoral Researcher at Delft University of Technology. From the spring of 2015, he worked as a senior researcher at Ericsson AB, Sweden. Since 2017, he has been a senior lecturer in the school of computing and communications, University of Technology Sydney, Australia. His current research interests include large scale distributed optimization and fundamental research of deep learning such as design of activation units and novel dropout techniques.