

Parallel Approaches for Data Mining in the Internet of Things Realm

Francesco Piccialli¹ · Salvatore Cuomo² ·
Gwanggil Jeon³

Published online: 17 March 2018

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

Recent studies show that 2.5 quintillion bytes of data per day are generated, and this is set to explode to 40 yottabytes by 2020. Much of this data is and will be generated from Internet of Things (IoT) devices and sensors. Billions of connected objects use the Internet every day capturing and producing data to be processed and excessive data is making great troubles to human beings. Data coming from IoT systems have a great diversity of types and therefore it becomes difficult to process by using state-of-the-art data processing techniques or traditional data processing platforms. In this scenario, the IoT realm requires more efficient and scalable data processing methods and, at the same time, raises additional challenges on data processing, mining and analytics. The first impression of the data produced by the IoT is the volume, a major challenge is how to process and cope with the massive amounts of rapidly scaling data deriving from sensors and devices. Collecting, storing and processing this data efficiently and quickly is vital to producing actionable, real-time insights. As it is well known, parallel and distributed computing have emerged in the last decades as well-developed research

✉ Gwanggil Jeon
gjeon@inu.ac.kr

Francesco Piccialli
francesco.piccialli@unina.it

Salvatore Cuomo
salvatore.cuomo@unina.it

¹ Department of Electrical Engineering and Information Technology, University of Naples FEDERICO II, Naples, Italy

² University of Naples “FEDERICO II”, Naples, Italy

³ Incheon National University, Incheon, Korea

areas in computer science and information technology. In this perspective, parallel and distributed computing techniques can be opportunely exploited in order to solve large-scale problems and process the data coming from the IoT paradigm.

This special issue is intended to provide a highly recognized international forum to presenting recent advances in parallel programming, data processing and the use of distributed or cloud systems in innovative paradigms like IoT in order to efficiently process and manage the huge amount of produced data. The ultimate objective is to bring together well-focused, top quality research contributions, providing to the general parallel programming community an opportunity to get an overall view of recent results, to identify the most promising avenues and to promote the visibility and relevance of data mining in the IoT. The intent is to raise collective awareness of the domain of parallel approaches for data mining in the IoT as a highly promising area to be pursued by the parallel programming research community.

The usage of computer vision adds a new paradigm in the field of animal biometric, and has recently received more attention due to the growing importance of identification and tracking of animal species or individual animals. Biometric characteristics help to develop a better representation and a better identification of different animal species and individual animals. The contribution by Sangaiah et al. “Group sparse representation approach for recognition of cattle on muzzle point images” proposes an effective approach for automatic cattle recognition based on the multiple features of muzzle points and the cattle face images. In this work, a comparative study among the well-established handcrafted texture feature extraction techniques and the appearance-based feature extraction techniques is presented. A detailed set of experimental results on muzzle point image database is also carried to prove the theory.

Infrared imaging has the advantage of all-weather working ability. Due to the limitation of the hardware and the high cost, the resolution of infrared image is very low. To improve the resolution of infrared images, this paper exploits super-resolution method for infrared images. The contribution by Wu et al. “Infrared image super-resolution with parallel random forest” proposes a super-resolution framework by using random forests. Existing methods adopts single regression model for SR while the single regression model tends to overfit training data and would lead to a poor performance. To solve this issue, authors adopt an ensemble regression model which is also known as random forests. In addition, a second order derivative filter is adopted, which can extract features on diagonal orientation.

Efficient and accurate vehicle detection has become one of challenging problems for complex urban traffic surveillance. As such, the contribution by Song et al. “Vehicle detection using spatial relationships GMM for complex urban surveillance in daytime and nighttime” proposes a new vehicle detection method using spatial relationship GMM for daytime and nighttime based on a high-resolution camera. In this work, the vehicle is treated as an object composed of multiple components, including the license plate, rear lamps and headlights. These components are localized using their distinctive color, texture, and region feature. Deriving plate color converting model, plate hypothesis score calculation and cascade plate refining were accomplished for license plate localization. Multi-threshold segmentation and connected component analysis are accomplished for rear lamps localization. Frame difference and geometric features similarity analysis are accomplished for headlights localization. After that,

the detected components are taken to construct the spatial relationship using GMM. Finally, similar probability measures of the model and the GMM, including GMM of plate and rear lamp, GMM of both rear lamps and GMM of both headlights are adopted to localize vehicle.

Target detection based on image/video as well as the change in the form of movement caused by camera imaging, algorithms are always designed complexly. However, object shelter and adhesion still cannot be fully resolved. Considering of that, the contribution by Song et al. “Target detection based on 3D multi-component model and inverse projection transformation” proposes a new method for target detection on true 3D space based on the inverse projection transformation and a mixing component model. Firstly, the inverse projective arrays parallel to target local surface are established on 3D space. Then, the 2D image is inversely projected to these planes through 3D point cloud re-projection, and a lot of inverse projective images with target local apparent characteristics are gained. After that, component HOG feature dictionaries are trained using the inverse projective images as samples, and on account of it, sparse decomposition approach is adopted to detect target local components. Finally, 3D centroid clustering for all the components is further used to identify the target.

The IoT based interoperable infrastructure is a convenient way for interaction and collaboration between students and teachers. As new learning styles develop, new tools and assessment methods are also needed. The contribution by Farhan et al. “A real-time data mining approach for interaction analytics assessment: IoT based student interaction framework” proposes a method which is to develop IoT based interaction framework and analysis of the student experience in electronic learning (eLearning) so that the students can take full advantage of the modern interaction technology and their learning can increase to a high level. IoT based infrastructure provides the facilities to fellow students about location awareness, fellows’ accessibility, social behavior and helping hand.

Currently, most of existing Galois/Counter Mode (GCM) architectures concentrated on power and area reduction but a compact and efficient hardware architecture should also be considered. The contribution by Paul et al. “High performance GCM architecture for the security of high speed network” proposes high-performance architecture for GCM. In order to achieve operating frequency and throughput, pipelined S-boxes are used in Advanced Encryption Standard (AES) algorithm. For a GCM realization of AES, a high-speed, high-throughput, parallel architecture is proposed. Simulations prove that the performance of the proposed work is around 17% higher than the existing architecture with 3 Gb/s throughput using TSMC 45-nm CMOS technology.

The capability for understanding data passes through the ability of producing an effective and fast classification of the information in a time frame that allows to keep and preserve the value of the information itself and its potential. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. A powerful tool is provided by Self-Organizing Maps (SOM). The goal of learning in the self-organizing map is to cause different parts of the network to respond similarly to certain input patterns. Because of its time complexity, often using this method is a critical challenge. In the contribution by Marcellino et al. “Parallel implementation of a machine learning algorithm on GPU,” authors propose a paral-

lel implementation for the SOM algorithm, using parallel processor architecture, as modern Graphics Processing Units (GPU) by CUDA.

The contribution by Yang et al. “Parallel heat kernel volume based local binary pattern on multi-orientation planes for face representation” proposes HKV–LBP–MOP approach. During the HKV–LBP–MOP, multi-scale heat kernel faces are captured in parallel and then reformulated as three-dimensional volume. Authors generate multi-orientation planes from the heat kernel volume, which reflects orientation co-occurrence statistics among different heat kernel faces. Finally, LBP analysis is applied on these multi-orientation planes of the heat kernel face volume to encode sufficient information for face representation. Simulations performed on ORL and Yale datasets reveal that the proposed approach is robust in varying lighting conditions, facial expressions, head pose, and facial occlusions. The proposed method is also more accurate than the current state-of-the-art face recognition methods.

The contribution by Mei et al. “Performance evaluation of GPU-accelerated spatial interpolation using radial basis functions for building explicit surfaces” focuses on evaluating the computational performance of parallel spatial interpolation with Radial Basis Functions (RBFs) that is developed by utilizing modern GPUs. The RBFs can be used in spatial interpolation to build explicit surfaces such as Discrete Elevation Models. When interpolating with large-size of data points and interpolated points for building explicit surfaces, the computational cost would be quite expensive. To improve the computational efficiency, authors specifically develop a parallel RBF spatial interpolation algorithm on many-core GPUs, and compare it with the parallel version implemented on multi-core CPUs. Five groups of experimental tests are conducted on two machines to evaluate the computational efficiency of the presented GPU-accelerated RBF spatial interpolation algorithm.

The purpose of abstractive summarization of multi-documents is to automatically produce a condensed version of the document text and maintain the significant information. Most of the graph-based extractive methods represent sentence as bag of words and utilize content similarity measure, which might fail to detect semantically equivalent redundant sentences. On other hand, graph based abstractive method depends on domain expert to build a semantic graph from manually created ontology, which requires time and effort. The contribution by Ahmad et al. “Abstractive text summarization based on improved semantic graph approach” presents a semantic graph approach with improved ranking algorithm for abstractive summarization of multi-documents. In this work, the semantic graph is built from the source documents in a manner that the graph nodes denote the predicate argument structures (PASs)—the semantic structure of sentence, which is automatically identified by using semantic role labeling (SRL); while graph edges represent similarity weight, which is computed from PASs semantic similarity.

We hope that this special issue would shed light on major developments in the area of parallel approaches for data mining in the IoT and attract attention by the scientific community to pursue further investigations leading to the rapid implementation of these technologies.

Acknowledgements We would like to express our appreciation to all the authors for their informative contributions and the reviewers for their support and constructive critiques in making this special issue possible.