A Research Roadmap of Big Data Clustering Algorithms for Future Internet of Things

Hind Bangui, Institute of Computer Science and Faculty of Informatics, Masaryk University, Brno, Czech Republic
Mouzhi Ge, Institute of Computer Science and Faculty of Informatics, Masaryk University, Brno, Czech Republic
Barbora Buhnova, Institute of Computer Science and Faculty of Informatics, Masaryk University, Brno, Czech Republic

ABSTRACT

Due to the massive data increase in different Internet of Things (IoT) domains such as healthcare IoT and Smart City IoT, Big Data technologies have been emerged as critical analytics tools for analyzing the IoT data. Among the Big Data technologies, data clustering is one of the essential approaches to process the IoT data. However, how to select a suitable clustering algorithm for IoT data is still unclear. Furthermore, since Big Data technology are still in its initial stage for different IoT domains, it is thus valuable to propose and structure the research challenges between Big Data and IoT. Therefore, this article starts by reviewing and comparing the data clustering algorithms that can be applied in IoT datasets, and then extends the discussions to a broader IoT context such as IoT dynamics and IoT mobile networks. Finally, this article identifies a set of research challenges that harvest a research roadmap for the Big Data research in IoT domains. The proposed research roadmap aims at bridging the research gaps between Big Data and various IoT contexts.

KEYWORDS

Big Data, Clustering Algorithm, Internet of Things, Machine Learning, Mobile Networks

1. INTRODUCTION

Internet of Things (IoT) is thanks to its increasing popularity becoming one of the key technological advancements of this decade. IoT is characterized by using smart and self-configuring objects that can interact with each other via global network infrastructure. This interaction among large number of heterogeneous objects makes IoT a disruptive technology for many industries, which enables ubiquitous and pervasive computing applications (Da et al., 2014; Ge et al., 2018). Accordingly, a wide range of industrial IoT applications have been developed and deployed in different domains such as energy management, transportation systems, agriculture, food processing, health monitoring, environmental monitoring, and security surveillance (Pourzolfaghar et al., 2018).

Since IoT advances the interconnectivity of the world around us, it plays an important role in the development of smart services (Pourzolfaghar and Helfert, 2017). Specifically,
the dynamic things collect different kinds of data from the real-world environment, and the extracted relevant information from the data can be used to improve and enrich our daily life with many kinds of context-aware applications (Dey, 2001), such as making cyber-physical systems of smart grid more robust and resilient (Wang et al., 2018), connecting patients and objects to one another for making their lives easier, constructing a modern model of hospital-centric care (Farahani et al., 2018), protecting location privacy to minimize potential hacking as well as unauthorized access to personal information (Sun et al., 2017).

IoT is an important source of contextual data with a large volume and fast velocity that contributes to the emergence to Big Data, which is typically defined based on three fundamental elements: volume (size of data), variety (different types of data from several sources) and velocity (data collected in real time). Moreover, additional characteristics to the 3V’s model have been introduced, such as (Manogaran et al., 2017) that adds value (benefits to various industrial and academic fields), veracity (uncertainty of data), validity (correct processing of the data), variability (context of data), viscosity (latency data transmission between the source and destination), virality (speed of the data send and receives from various sources) and visualization (interpretation of data is more concerned and identification of the most relevant information for the users). Yet, the 3V’s model is considered to be the basis of the Big Data concept (Kitchin, 2017).

The fusion of Big Data and IoT technologies has created opportunities for the development of services within smart environments like Smart Cities. To support the services with information, Big Data analytics for IoT have emerged to process the data collected from different sources in the smart environment (Chen et al., 2016). On one hand, Big Data and its technologies have opened new opportunities for industries and academics to develop new IoT solutions. On the other hand, the advancement of IoT is producing vast amount and different types of data, especially after the appearance of the emerging 5G (Mavromoustakis et al., 2016), which is becoming increasingly difficult to process. Therefore, the fusion of Big Data and IoT, as well as the highly dynamic evolution of the two domains, create new research challenges, which have so far not been addressed by the research community.

One of the key and critical operations of Big Data processing and analytics is the clustering, currently supported by a number of algorithms. Previously, we have reviewed the advantages and disadvantages of clustering algorithms, which indicate that clustering is one of the key factors to supply the fusion of Big Data, cloud computing, mobile environment and IoT technologies (Bangui et al., 2018).

This paper is an extended version of (Bangui et al., 2018), the key findings from the review of the clustering algorithms in Big Data are presented together with the illustration of how clustering algorithms in Big Data can be used in IoT. Based on the review, the paper presents a set of research challenges to clarify the research gaps between Big Data and IoT. In addition to (Bangui et al., 2018), this paper has included the comparisons of different clustering algorithms for IoT data, re-structured the research challenges and developed a research roadmap for Big Data research in IoT domains. The whole paper text is largely extended, re-organized and further polished.

The remainder of the paper is organized as follows. Section 2 carries out a systematic study on clustering algorithms in the Big IoT Data, which includes clustering algorithm characteristics and classifications, summarizing the features of IoT data, as well as comparing the clustering algorithms for IoT data. Based on the literature review and experimental results, Section 3 presents a research roadmap that consists of major challenges and opportunities related to the fusion of IoT, cloud computing, mobile environment and Big Data technologies in the 5G networks. Finally, section 4 concludes the work and outlines the future research.
Optimal Location of the Workpiece in a PKM-based Machining Robotic Cell
www.igi-global.com/chapter/optimal-location-workpiece-pkm-based/72830?camid=4v1a

Particle Swarm Optimization Algorithms Inspired by Immunity-Clonal Mechanism and Their Applications to Spam Detection
Ying Tan (2012). Innovations and Developments of Swarm Intelligence Applications (pp. 182-205).
www.igi-global.com/chapter/particle-swarm-optimization-algorithms-inspired/65813?camid=4v1a