Chapter 75
Efficiently Processing Big Data in Real-Time Employing Deep Learning Algorithms

Murad Khan

Sarhad University of Science and Information Technology, Pakistan

Bhagya Nathali Silva

Kyungpook National University, South Korea

Kijun Han

Kyungpook National University, South Korea

ABSTRACT

Big Data and deep computation are among the buzzwords in the present sophisticated digital world. Big Data has emerged with the expeditious growth of digital data. This chapter addresses the problem of employing deep learning algorithms in Big Data analytics. Unlike the traditional algorithms, this chapter comes up with various solutions to employ advanced deep learning mechanisms with less complexity and finally present a generic solution. The deep learning algorithms require less time to process the big amount of data based on different contexts. However, collecting the accurate feature and classifying the context into patterns using neural networks algorithms require high time and complexity. Therefore, using deep learning algorithms in integration with neural networks can bring optimize solutions. Consequently, the aim of this chapter is to provide an overview of how the advance deep learning algorithms can be used to solve various existing challenges in Big Data analytics.

INTRODUCTION

Big Data and deep computing have gained a smashing popularity over the past few decades. The emergence of Big Data was accompanied by the exponential growth of digital data. Big Data has been defined in multiple aspects and perspectives. In general, Big Data is a prodigious amount of digital data, which is strenuous to manage and analyze using generic software tools and techniques. According to the National

Security Agency, 1826 petabytes are processing by the internet on a daily basis ("National Security Agencey: Missions Authorities, Oversight and Partnerships", 2013). Surprisingly, in 2011 it was found that the world's data volume has grown in nine times within five years. This extraordinary growth rate is estimated to reach 35 trillion gigabytes in 2020 (Gantz & Reinsel, 2011). Due to this exponential digital data generation, Big Data continues to receive an extreme attention from the industrial experts as well as from the interested researchers. In fact, Big Data requires expeditious processing over voluminous data sets with high variety and high veracity (Zhang, Yang and Chen, 2016). Therefore, it creates a compelling demand to discover and to adopt technologies capable of speedy processing on heterogeneous data. Numerous embedded devices connected to the network generates heterogeneous data. Figure 1 illustrates a classical architecture of heterogeneous devices connected over different communication technologies. Generic characteristics of Big Data is known as three V's of Big Data i.e. variety, velocity, and volume. A variety of data is referred to multiple formats of data being stored. For example, a collection of text, image, audio, video, numeric, etc. data types in structured, semi-structured, and unstructured forms are considered as data sets with variety. The size aspect of the data is the volume of data. In a modern technological era, data volume is rapidly growing with the invention of social media and popularity of embedded devices. In fact, the data generation speed is known as the velocity of Big Data. The technological advancements have influenced the dramatic increase in the velocity. Moreover, Big Data includes incomplete, redundant data as well as inaccurate and obsolete data. Thus, it is denoted that Big Data consists of high veracity data. Consequent to the rapid growth in digital data, myriads of opportunities in numerous fields i.e. educational services, enterprises, manufacturing services, social networking and much more are emerging. Consequently, these opportunities of Big Data have geared the research community towards data-driven discovery. Indeed, Big Data phenomenon has influenced all aspects of the social lifestyles in the modern world. Even though, Big Data occupies a colossal amount of data, discovering precise knowledge from the gathered data is not an easy task. Henceforth, the spotlight is focused towards the standard representation, storage, analysis, and mining of Big Data. The extreme heterogeneous nature of Big Data hinders the capability of feature learning by conventional data mining methods and algorithms. In fact, not only the advancements of the existing technologies, it is required collaboration among associative teams as well in order to extract valuable knowledge from Big Data. The advancements of the computation capabilities and enhanced machine learning mechanisms have broadened boundaries of data analytics and knowledge discovery procedures. However, the extraction of knowledge from Big Data using conventional learning algorithms is an extremely laborious task. It might be impossible in certain scenarios. The large volumes of data demand the scalability of the algorithm in order to process Big Data. Moreover, the great variety of data requires the algorithm to identify hidden relationships among heterogeneous data. The interest groups have yearned to discover novel mechanisms to accomplish the tedious task, which cannot be fulfilled by the conventional learning algorithms. Efforts in the recent past have identified that the integration of deep learning with high-performance computation offers favorable outcomes in knowledge extraction from Big Data.

The conventional learning algorithms occupy shallow learning architectures. In contrast, the deep learning algorithms incorporate deep architectures (Bengio and Bengio, 2000). The deep learning mechanisms utilize machine-learning techniques such as supervised learning and unsupervised learning to learn hierarchical relationships among heterogeneous data. Accordingly, it classifies the data autonomously (Ranzato, Boureau and Yann, 2007). Deep learning algorithms have gained a wider acceptance from the research community due to its remarkable performance in various domains i.e. speech recognition, computer vision, fraud detection, and advanced search. The deep learning algorithms classify and process

Related Content

Emerging Missing Data Estimation Problems: Heteroskedasticity; Dynamic Programming and Impact of Missing Data
Tshilidzi Marwala (2009). *Computational Intelligence for Missing Data Imputation, Estimation, and Management: Knowledge Optimization Techniques (pp. 276-301).*
www.irma-international.org/chapter/emerging-missing-data-estimation-problems/6805

Convolutional Approach Also Benefits Traditional Face Pattern Recognition Algorithm [208!]
Yunke Li, Hongyuan Shi, Liang Chenand Fan Jiang (2019). *International Journal of Software Science and Computational Intelligence (pp. 1-16).*
www.irma-international.org/article/convolutional-approach-also-benefits-traditional-face-pattern-recognition-algorithm-208/247132

Recurrent Neural Network (RNN) to Analyse Mental Behaviour in Social Media
Hadj Ahmed Bouarara (2021). *International Journal of Software Science and Computational Intelligence (pp. 1-11).*
www.irma-international.org/article/recurrent-neural-network-rnn-to-analyse-mental-behaviour-in-social-media/280513

Cost-Sensitive Learning in Medicine
Alberto Freitas, Pavel Brazdiland Altamiro Costa-Pereira (2012). *Machine Learning: Concepts, Methodologies, Tools and Applications (pp. 1625-1641).*
www.irma-international.org/chapter/cost-sensitive-learning-medicine/56217

Remote Conversation Support for People with Aphasia
Nilar Aye, Takuro Ito, Fumio Hattori, Kazuhiro Kuwabaraand Kiyoshi Yasuda (2010). *International Journal of Software Science and Computational Intelligence (pp. 72-85).*
www.irma-international.org/article/remote-conversation-support-people-aphasia/39106