Deep Learning With Analytics on Edge

Kavita Srivastava
Institute of Information Technology and Management, GGSIP University, India

EXECUTIVE SUMMARY

The steep rise in autonomous systems and the internet of things in recent years has influenced the way in which computation has performed. With built-in AI (artificial intelligence) in IoT and cyber-physical systems, the need for high-performance computing has emerged. Cloud computing is no longer sufficient for the sensor-driven systems which continuously keep on collecting data from the environment. The sensor-based systems such as autonomous vehicles require analysis of data and predictions in real-time which is not possible only with the centralized cloud. This scenario has given rise to a new computing paradigm called edge computing. Edge computing requires the storage of data, analysis, and prediction performed on the network edge as opposed to a cloud server thereby enabling quick response and less storage overhead. The intelligence at the edge can be obtained through deep learning. This chapter contains information about various deep learning frameworks, hardware, and systems for edge computing and examples of deep neural network training using the Caffe 2 framework.

INTRODUCTION

Deep Learning is a subset of Machine Learning which is being used widely in many applications related to Computer Vision (CV) and Speech Processing. There are several techniques that belong to Deep Learning. These techniques include Deep

Copyright © 2021, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

Neural Network (DNN), Convolution Neural Network (CNN), Recurrent Neural Networks (RNN), Long Short Term Memory (LSTM) and Transfer Learning.

All of the deep learning methods have similar characteristics. That is, these methods are data hungry. They perform better with more data. These methods require high computation power and need longer time for training and inferences.

Since deep learning methods are resource intensive in terms of both computing power and storage requirement they often need high performance of cloud server. However, the enabling applications of deep learning such as autonomous vehicles and self-driving cars, home automation and security systems, face detection applications and speech recognition systems require quick response which is not possible when the analysis is done on the cloud server because of latency inherent with cloud processing. Another problem associated with the analysis done on cloud server is that network connectivity is not available all the time.

Addressing of all these issues require the analysis and computation part to be done locally at the network edge. With data analysis and predictions done near the location of data collection, the response time can be reduced substantially. This scenario leads to the emergence of a new computing paradigm called Edge Computing.

Edge computing is distributed in nature as opposed to the Cloud Computing which makes use of a centralized cloud server. Edge computing is mostly applicable to Autonomous Systems (AS), Cyber-Physical Systems and Internet of Things (IoT) applications.

IoT applications comprise of an embedded system, communication system and several sensors. Sensor nodes don't either need extensive computing power offered by the cloud server or the storage space offered by cloud. The concept of Edge Computing refers that the computation is performed in close proximity to the end user. It means the computation is either performed locally on the sensor nodes or on a server near to these nodes, that is, at the network edge.

Edge computing offers a number of benefits to the end user. Edge computing preserves the privacy of personal data of users. The user data need not be sent to the cloud server for training of model. Only the model information is transferred to the cloud server. Since bulk data is not transferred, less number of network resources are required. The edge computing provides scalability as more edge devices can be added easily. As shown in Figure intelligent IoT and other applications can utilize Edge Intelligence. The pre-trained Neural Network Model is deployed at the network edge whereas the model training is performed the backend cloud server.

In the rest of this chapter, state-of-the-art literature survey is provided in section 2. Section 3 describes several application use cases which require the usage of deep learning along with computation on an edge device. Section 4 describes hardware systems and platforms that run deep learning applications. Section 5 provides a comprehensive discussion on various Deep Learning Frameworks both for training

21 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/deep-learning-with-analytics-on-edge/271708

Related Content

Learning Bayesian Networks
Marco F. Ramoniand Paola Sebastiani (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1124-1128).*
www.irma-international.org/chapter/learning-bayesian-networks/10962

Clustering Analysis of Data with High Dimensionality
Athman Bouguettayaand Qi Yu (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 237-245).*
www.irma-international.org/chapter/clustering-analysis-data-high-dimensionality/10827

Secure Computation for Privacy Preserving Data Mining
Yehuda Lindell (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 1747-1752).*
www.irma-international.org/chapter/secure-computation-privacy-preserving-data/11054

Cluster Analysis in Fitting Mixtures of Curves
Tom Burr (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 219-224).*
www.irma-international.org/chapter/cluster-analysis-fitting-mixtures-curves/10824

Feature Extraction/Selection in High-Dimensional Spectral Data
Seoung Bum Kim (2009). *Encyclopedia of Data Warehousing and Mining, Second Edition (pp. 863-869).*
www.irma-international.org/chapter/feature-extraction-selection-high-dimensional/10921