

Chapter 4

Deep Learning–Enabled Edge Computing and IoT

Amuthan Nallathambi

 <https://orcid.org/0000-0001-8830-3353>

AMC Engineering College, Visvesvaraya Technological University, India

Kannan Nova

Microsoft, USA

ABSTRACT

Deep learning is a new approach to artificial intelligence that enables edge-computing systems to learn from data and take decisions without human intervention. Edge computing is a technique for coping with the increasing demand for streaming data. This is especially important in the case of applications that involve computationally intensive tasks such as driverless cars, autonomous drones, and smart cities. Edge computing is the provision of computing, big data analytics, and storage in such a way that the data comes to the processing power and not vice versa. It relies on a decentralized approach where computational resources are provided at the edge of networks. Edge computing is an emerging field that's getting attention from many vendors and researchers. The data generated by IoT devices is usually too large and complex for cloud-based storage and processing. That's why edge computing can handle data at the source of generation in real time, which speeds up the process of decision making.

INTRODUCTION

In this chapter, we explore how deep learning enables edge computing applications at scale by predicting user preferences. Deep learning is an artificial intelligence algorithm which can learn from example data and improve its predictions. Deep learning enables edge computing applications at scale with real-time analytics. Deep learning is a subset of machine learning that use algorithms based on artificial neural networks to learn abstract concepts from data. Deep learning models have an architecture composed of many layers or modules, each designed to process and extract data from its inputs and pass it through multiple processing steps to generate outputs.

DOI: 10.4018/978-1-6684-6275-1.ch004

Deep learning models can be identified as following based on the number of layers in their architecture:

Single-layer networks: These models are composed of one layer with only one processing step. One layer includes a set of inputs and a single output. A single-layer network is an example of a perceptron model.

Multi-layer networks: These models have two or more layers in the architecture, each performing different tasks. The multiple hidden layers are introduced between input and output layers depends upon the optimization or hyper parameters. This can be achieved by means of a feed-forward architecture, whereby the output of one layer is input to the next layer, or a recurrent architecture with feedback. A multilayer network is an example of a deep neural network. The following diagrams illustrate three different types of multilayer networks: Feedforward multi-layer perceptron model: This is the simplest form of a multilayer network.

Recurrent neural network model: The recurrent neural network (RNN) model is an artificial intelligence technique used to process sequential data and produce a sequence of outputs where each successive value depends on the values of previous entries. This is an alternative to the backpropagation network. This is a more sophisticated form of the feedforward architecture. They are trained using the standard batch training, but the time required for learning is reduced significantly compared to standard neural networks. This type of model is commonly used in machine learning to process time-based data. For instance, this kind of model can be used to predict the next word of a text or the next impulse (e.g., sound, gesture) that an actor makes in a video sequence.

Convolution Neural Network: A popular deep learning algorithm is called a Convolutional Neural Network. The Convolutional Neural Network works by using a lot of “neurons” and they work together to solve problems. It is used in image classification, image recognition, object detections, and autonomous self-driving cars.

Deep neural network model: This most complex and powerful type includes many layers, typically with multiple hidden or intermediate layers in each layer. As illustrated in the figure.1,

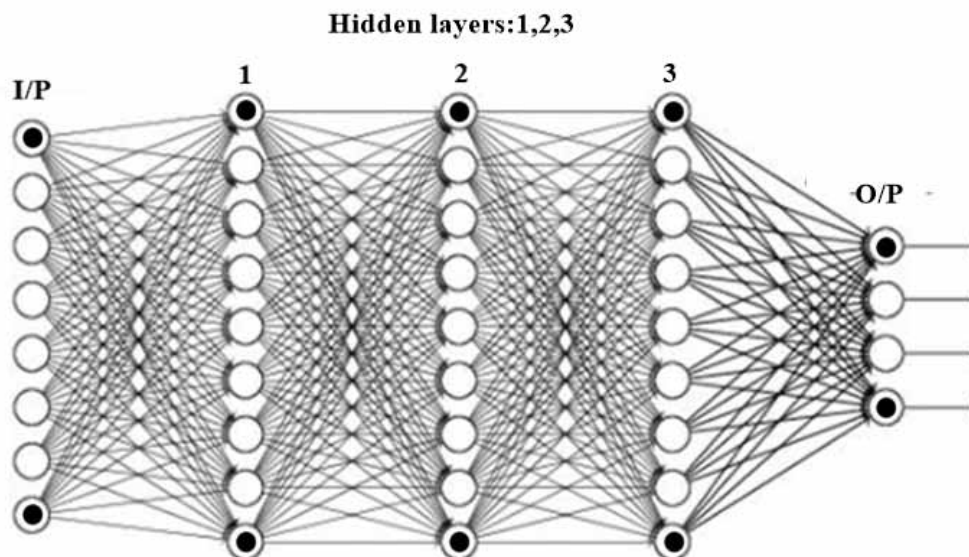


Figure 1. Deep neural network model

23 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-enabled-edge-computing-and-
iot/316015](http://www.igi-global.com/chapter/deep-learning-enabled-edge-computing-and-iot/316015)

Related Content

Mobile World Wide Web Content

Wen-Chen Hu (2009). *Internet-Enabled Handheld Devices, Computing, and Programming: Mobile Commerce and Personal Data Applications* (pp. 139-161).

www.irma-international.org/chapter/mobile-world-wide-web-content/24701

OO-H Method: Extending UML to Model Web Interfaces

Jaime Gomezand Cristina Cachero (2003). *Information Modeling for Internet Applications* (pp. 144-173).

www.irma-international.org/chapter/method-extending-uml-model-web/22972

Cultural Tourism and the Tourist Experience in the Digital Era

Eunice Ramos Lopes, Paulo Alexandre Santosand João Tomaz Simões (2022). *Handbook of Research on Digital Communications, Internet of Things, and the Future of Cultural Tourism* (pp. 1-18).

www.irma-international.org/chapter/cultural-tourism-and-the-tourist-experience-in-the-digital-era/295494

The Future of Digital Tourism Alternatives in Virtual Reality

Zuleyhan Baranand Huseyin Baran (2022). *Handbook of Research on Digital Communications, Internet of Things, and the Future of Cultural Tourism* (pp. 58-84).

www.irma-international.org/chapter/the-future-of-digital-tourism-alternatives-in-virtual-reality/295497

A Neural Network-Based Automatic Crop Monitoring Robot for Agriculture

E. Udayakumar, S. Balamuruganand P. Vetrivelan (2019). *The IoT and the Next Revolutions Automating the World* (pp. 203-212).

www.irma-international.org/chapter/a-neural-network-based-automatic-crop-monitoring-robot-for-agriculture/234031