Chapter 5
Distributed Deep Learning for IoT

Amuthan Nallathambi
https://orcid.org/0000-0001-8830-3353
AMC Engineering College, Bengaluru, India

Sivakumar N.
Department of Mechanical Engineering, India

Velrajkumar P.
CMR Institute of Technology, Visvesvaraya Technological University, India

ABSTRACT

Distributed deep learning is a type of machine learning that uses neural networks to learn and make predictions at scale. This is achieved by having many different computer systems that are connected via the internet. This allows for more parallel processing and faster results. In addition, when it comes to IoT, this type of technology can be used in conjunction with sensors and other devices to create more accurate predictions about the environment around us. Distributed deep learning can be used in many ways with the IoT because it can be applied to various aspects of IoT data processing, such as image recognition, speech recognition, natural language processing (NLP), or anomaly detection. The neural net is the most computationally intensive component of the system, and it requires a significant amount of energy. To make this system more cost-effective, there are two ways to lower the number of memory accesses: by reducing the size of images (so precision decreases), or by increasing network bandwidth so that there are fewer loop iterations required for each memory access.

INTRODUCTION

Deep neural networks are able to recognize objects in an image through pattern recognition and learn from these patterns to recognize text in a piece of text. Because of the way deep neural networks operate, they are able to train themselves with minimal supervision. This makes them a popular choice for machine learning and artificial intelligence applications because they can learn on their own rather than

DOI: 10.4018/978-1-6684-6275-1.ch005

be programmed explicitly (Wikimedia Foundation, 2022, July 9). A few notable examples of deep neural networks in action are Google Translate, AlphaGo, and the Microsoft Tay Twitter bot.

In the typical method of training for deep learning, users first upload a local dataset to a cloud centre, where it is then trained using the huge computer capabilities of the cloud centre. The uploading of a local dataset to a centralized cloud centre that is controlled by a third party puts the user's data privacy at risk. Additionally, the uploading of multimedia data will use up a lot of mobile users' bandwidth as well as the storage space of the cloud centre, making it difficult to add more edge devices. We present an edge-enabled distributed deep learning platform that tackles these two difficulties by dividing a conventional deep learning training network into a front subnetwork and a rear subnetwork. This allows the platform to make use of distributed computing resources. In particular, the front subnetwork, which is comprised of multiple layers, is established in close proximity to the data source and trained independently at each edge device using the information that is local to that device. In order for the back subnetwork to be trained at a later time, the outputs of all of the front subnetworks are transferred to the back subnetwork, which is installed in a cloud data centre (Sun, Q., 2021).

Because edge devices do not send their original data to the cloud centre, the platform safeguards the data's confidentiality while still allowing for rapid expansion. On top of that, there are two more steps taken to protect data privacy: asymmetric encryption technology is used to make sure that parameters sent between edge servers and the cloud centre are safe and complete, and blockchain technology is used to track what people do on this platform and build trust among them. Both of these technologies are used to ensure that parameters sent between edge servers and the cloud centre are safe and complete.

In neural networks, there are a million parameters that make up the model, and learning these parameters requires a massive amount of data. This approach is time-consuming and computationally costly. A new approach to artificial neural networks is tackling one of the main problems in AI: training and deployment. Distributed Deep Learning (DDL) uses a simple technique that can potentially reduce the typical AI development time by half., Distributed deep learning allows for parallel experiments over many devices in order to reduce training time. This is accomplished with the use of remote GPUs, software containers, and cloud computing services. Using this, the company is able to train deep learning models with a single CPU in just a few hours. As well as this, it's possible to train models with many CPU cores as the power of two is still being used.

The future of artificial intelligence will likely be powered by DDL, which will help improve the usability, performance, and accuracy of the systems. DDL is a data-driven decision-making tool that helps in the execution of business strategies by providing algorithms and insights that can be used as models. These models are capable of predicting outcomes, identifying high risks, and optimizing performance.

The following are the features of Decision Lens:

- Provides insights and actionable information to help companies make better decisions.
- Challenges and opportunities based on unique business challenges -Predicts outcomes (both positive and negative) in a variety of industries.
- Helps identify high-risk areas that can improve performance by highlighting promising

Artificial intelligence (AI) has been the topic of a lot of speculation and debate in recent years. The future of AI is bright as scientists continue to tinker with the technology. With advancements in machine learning, machine learning algorithms will be used to power the future of AI and improve its usability, performance, and accuracy as well. As AI becomes more widespread, the need for a standardized frame-