Chapter 1
Deep Learning for Big Data Analytics

Priti Srinivas Sajja
Sardar Patel University, India

Rajendra Akerkar
Western Norway Research Institute, Norway

ABSTRACT

Traditional approaches like artificial neural networks, in spite of their intelligent support such as learning from large amount of data, are not useful for big data analytics for many reasons. The chapter discusses the difficulties while analyzing big data and introduces deep learning as a solution. This chapter discusses various deep learning techniques and models for big data analytics. The chapter presents necessary fundamentals of an artificial neural network, deep learning, and big data analytics. Different deep models such as autoencoders, deep belief nets, convolutional neural networks, recurrent neural networks, reinforcement learning neural networks, multi model approach, parallelization, and cognitive computing are discussed here, with the latest research and applications. The chapter concludes with discussion on future research and application areas.

INTRODUCTION

Deep learning refers to a kind of machine learning techniques in which several stages of non-linear information processing in hierarchical architectures are utilized for pattern classification and for feature learning. Recently, it also involves a hierarchy of features or concepts where higher-level concepts are defined from lower-level ones and where the same lower-level concepts help to define higher-level ones. With the enormous amount of data available today, big data brings new opportunities for various sectors; in contrast, it also presents exceptional challenges to utilize data. Here deep learning plays a key role in providing big data analytics solutions. The chapter discusses in brief fundamentals of big data analytics, neural network, deep learning. Further, models of deep learning are analyzed with their issues and limitations along with possible applications. Summary of the literature review is also provided in this chapter. Further, future possible enhancements are also listed in the domain. This chapter is organized as follows.

DOI: 10.4018/978-1-5225-5852-1.ch001

Section 1 introduces various fundamental topics such as big data analytics, artificial neural network, and deep learning. Section 2 highlights big data analytics by discussing large scale optimization, high dimensional data handling, and handling dynamic data. Section 3 discusses different deep models such as autoencoders, deep belief nets, deep convolutional neural networks, recurrent neural networks, reinforcement learning neural networks, multi model approach, parallelization, and cognitive computing with latest research and applications. Section 4 discusses some successful applications of deep learning for big data analytics. Section 5 discusses the issues and problems with the deep learning. Section 6 concludes the paper with summary and provides discussion on the work done so far and future research and application areas in the domain.

Big Data Analytics

Data word came from 'Datumn' which means 'thing given'. It has been used since early 1500 i.e. since beginning of computing. With the evolution of computing technology, the word has become more popular. Data are raw observations from domain of interest. It can be collection of numbers, words, measurements, or textual description of things. Obviously, data is everywhere and serves as an important base for business related decision making. It is also said that data is currency of knowledge as it provides basis of reasoning and analysis. Every business generates lots of data, which further act as a good resource to analyze, understand and improve the business. It is really an irony that the data which can help in improving quality of business makes the life miserable just because of our limitations to understand and use it properly. Such data creates a big problem due to its size, unstructuredness and redundancy. Some researchers identify the parameters like volume, velocity and variety as main reasons of the problem to handle data. According to Eric Horvitz and Tom Mitchell (2010) and James Manyika et al., (2011) such data when analyzed and used properly, offers a chance to solve problems, accelerates economic growth, and improve quality of life.

Big data is also a kind of data but big enough to handle. For example, consider a medical store having cough syrup bottles on shelves. The labels on the bottles show medicine name, type, components, batch number and expiry date. These data are very structured, small in amount (even for thousands of bottles in many batches, and such many medicines in the store) and homogeneous in nature. On the other hand, why the particular medicine expires so early requires really a big amount of data (and big amount of time and effort also). Such big data can act as a good resource to increase productivity and hence improves businesses in terms of quality, brand image and customer surplus. Effective use of big data can be one of the key factors of competition and growth for individual firm. It is predicted that by 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions (Manyika, et al., 2011).

Big data handling encompasses activities in five dimensions. The first step is the data procurement, where data are identified and procured as per the requirement and available sources. Prior to the data collection, it is required to have proper vision and plan to make sure that how these data will be handled and what are the expectations. The volume, velocity and variety of the data make these procedures more challenging. Since the data is big, procurement would not be an easy job. It may leave some gap and results in missing data. To compensate the gap, data processing is needed, which is the second step. One may use soft computing techniques (such as fuzzy logic) to predict the missing data and fill in the gap. Next step is to find out a proper representation scheme so that the complete set of data can be presented