Chapter 35

Novel Scalable Deep Learning Approaches for Big Data Analytics Applied to ECG Processing

Rostom Mennour

Constantine 2 University, Algeria

Mohamed Batouche

Constantine 2 University, Algeria

ABSTRACT

Big data analytics and deep learning are nowadays two of the most active research areas in computer science. As the data is becoming bigger and bigger, deep learning has a very important role to play in data analytics, and big data technologies will give it huge opportunities for different sectors. Deep learning brings new challenges especially when it comes to large amounts of data, the volume of datasets has to be processed and managed, also data in various applications come in a streaming way and deep learning approaches have to deal with this kind of applications. In this paper, the authors propose two novel approaches for discriminative deep learning, namely LS-DSN, and StreamDSN that are inspired from the deep stacking network algorithm. Two versions of the gradient descent algorithm were used to train the proposed algorithms. The experiment results have shown that the algorithms gave satisfying accuracy results and scale well when the size of data increases. In addition, StreamDSN algorithm have been applied to classify beats of ECG signals and provided good promising results.

INTRODUCTION

In the recent few years, big data has been one of the most attractive fields for both industry and academy. This is due to the vast opportunities that big data can give to several sectors in real life. Big data refers to the amounts of data that exceed the storage and computation capacities provided by traditional databases and data processing engines. The size of data has considerably grown in the last decade, with the

emergence of social networks, internet of things, cloud computing and other technologies that have led to what is called the data explosion. In 2011, the size of data had grown nine times in just five years (Gantz & Reinsel, 2011), and will soon (by 2020) reach 35 zettabytes (10^{21} bytes) (Gantz & Reinsel, 2010).

Those massive amounts of data generated continuously at unprecedented scale have to be exploited. Machine learning and data mining techniques have been used for many years to extract useful information from complex data generated in several fields like bioinformatics, healthcare, economy, social media, and so on. Now, it is harder to apply those techniques in the actual conditions. As mentioned before, the size of data has considerably grown, and the conventional learning methods were not designed to deal with this kind of phenomenon. It is imperative to rethink the way we develop our models and algorithms so that they will be able to meet the requirements of big data. This will considerably improve the quality of life, and bring big advances in science and engineering.

Recently, a new learning technique has emerged. In the machine learning field, deep learning is one of the most popular research areas nowadays. It uses deep hierarchical architectures inspired from the neocortex in the human's brain, to learn from large amounts of supervised and/or unsupervised data. These techniques have won the most important places in the world of machine learning, and have managed to achieve the best performance rankings becoming the state of the art in this field (Chen & Lin, 2014). Due to the high performances of the deep learning algorithms compared to classic ones, more and more academicians have turned to them for diverse applications like image processing, speech recognition, text and language analytics, bioinformatics and drug discovery and even more (Min, Lee, & Yoon, 2016). For instance, Google's DeepMinde (Hassabis, Suleyman, & Legg, 2014) has created the Deep Q-networks algorithm that can play Atari games (Mnih, Kavukcuoglu, Silver, Graves, & Antonoglou, 2013) successfully. The same company has also created a program (AlphaGo), based on deep learning, that has, for the first time, won a Go game versus a professional human player.

One of the most challenging issues for machine learning, especially for deep learning, in the era of big data, is the inability to use all of the data available (Najafabadi, Villanustre, Khoshgoftaar, Seliya, Wald, & Muharemagic, 2015). One good solution is to scale up those algorithms and use distributed and parallel paradigms to design them. Engines that are designed for processing massive datasets, like MapReduce (Dean & Ghemawa, 2008) and Spark (Zaharia, Chowdhury, Franklin, Shenker, & Stoica, 2010) are very suitable for this task. Their high ability to scale to large clusters of machines and to tolerate faults when the program is running makes them the best way to face the data size challenge. Also, the algorithms have to be parallelizable and fit well with these models. Given present circumstances, this task is considered to be crucial.

Another big challenge for deep learning in the big data context is velocity; data arrive at high speed to be processed on time (Najafabadi, Villanustre, Khoshgoftaar, Seliya, Wald, & Muharemagic, 2015). The fact is, current machines cannot hold all the data in memory, especial when it comes to big data applications. The solution for this problem consists of online learning, where one instance is learned at a time. Only a few works have been done in this context (Chen & Lin, 2014) and it's imperative to adapt deep learning to handle big streaming data, as the need for such approaches is becoming very significant.

On the other hand, Clinical databases have accumulated huge quantities of information about patients and their medical conditions. The danger of heart diseases requires an extremely fast analysis. A diagnosis at the right time can save a patient's life, while a delay could prove to be very dangerous. The health information of a patient, especially those related to his heart, are most relevant and valuable at the time of their arrival, that's why on the fly predictions must be done immediately. Also, it is time for the healthcare industry to move from the era of analyzing our health history to the age of managing its

19 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/novel-scalable-deep-learning-approaches-for-
big-data-analytics-applied-to-ecg-processing/237896

Related Content

Dynamic Monitoring of Forest Volumes by a Feature Extraction Method
Xu Jieand Dawei Qi (2018). *International Journal of Software Science and Computational Intelligence (pp. 53-68).*
www.irma-international.org/article/dynamic-monitoring-of-forest-volumes-by-a-feature-extraction-method/207745

Challenges in the Design of Adoptive, Intelligent and Cognitive Systems
W. Kinsner (2009). *International Journal of Software Science and Computational Intelligence (pp. 16-35).*
www.irma-international.org/article/challenges-design-adoptive-intelligent-cognitive/34086

Granular Computing and Human-Centricity in Computational Intelligence
Witold Pedrycz (2010). *International Journal of Software Science and Computational Intelligence (pp. 16-31).*
www.irma-international.org/article/granular-computing-human-centricity-computational/49129

A Proficient Hybrid Framework for Image Retrieval
Rajkumar Soundrapandiyanand Ramani Selvanambi (2021). *Applications of Artificial Intelligence for Smart Technology (pp. 206-219).*
www.irma-international.org/chapter/a-proficient-hybrid-framework-for-image-retrieval/265587

A High Level Model of a Conscious Embodied Agent
Jirí Wiedermann (2010). *International Journal of Software Science and Computational Intelligence (pp. 62-78).*
www.irma-international.org/article/high-level-model-conscious-embodied/46147