An Insight Into Deep Learning Architectures

Nishu Garg
VIT University, India

Nikhitha P
VIT University, India

B. K. Tripathy
VIT University, India

INTRODUCTION

For molding the world to incredible amplitude, appropriate learning of computers has inevitable participation. This ability of the computer of deploying the world can be referred as intelligence. For making a machine intelligent, it has to go through a series of training processes where the information is stored in an organized way making it easier to relate the so gathered information with the real life scenarios. Storing all the information in the machine manually is a troublesome task, especially when the information is abundant like in the case of sophisticated artificial intelligence tasks. This is why the learning algorithms have gained attention among researches to store huge information at a stretch. Many learning algorithms are established and capable of understanding the view, but failed to express efficiently in natural language. Semantic understanding of these algorithms was restricted by some degree which is the necessity and should be expanded. So it was figured out that the prior algorithms were incapable of maintaining the interaction towards the humans through various semantic and visual mediums. To resolve this problem of AI task, Deep architectures were introduced, so that the machines can be trained and can be made efficient in the areas they are lacking.

Information retrieval have widely classified into two subdivisions: image retrieval and the text retrieval (A.G. Abby, 2000). Both approaches were emerged in the mid 60-70's. In order to carry out the several tasks like indexing, searching and retrieval data has to be well organized. In this work, the focus is on the image retrieval perspective. Image retrieval is divided into query or text based and content based approach (P.A. Vikhar, 2010). Our investigation is centralized around the first methodology.Text based image retrieval is basic and fundamental one which makes use of searching through simple query word (T.Westrveld,, J.C Gemert,., R. Cornacchia,, D Hiemstra., A. P. Vries 2005). Earlier work to this text based approach was database management systems which binds the images with text. Multi-dimensional indexing, data modeling and query evaluation are the exemplary research lines. But the fore mentioned procedures turned out ineffectual for the complex data link structures, language pairs, sophisticated image data characteristics, question answering, annotation and context retrieval. These made a way to design a new approach that addresses the various issues.

The basic question found out while natural language processing is " Is there exists any analog program loaded in computer which converts a portion of textual data represented in English into a

DOI: 10.4018/978-1-5225-2255-3.ch393

computer understandable data model that clearly depict the actual text meaning?". While taking care of this problem, computer specialists have to get down for fewer objectives that represent lesser aspects of text data. Natural language processing is a flexible architecture which focuses on learning representations useful for the specific tasks. The NLP had evolved in a way for seeking a model which converts human understandable language (English in the beginning) into a machine friendly data representation that conveys the meaning of text without any ambiguity. The earlier times the evolutionary idea behind the NLP seemed to be very broad and vast for the researchers so they settled down for lesser goals of representing limited textual information by extracting simpler representations. Text processing in natural language processing encompasses by bag-of-words, removal of frequent words from the input text query (stop words), identifying the root words (stemming), spell check, text and controlled vocabulary mapping and named entity recognition.

BACKGROUND

The concept of neural networks derived from the human nervous system by the way of answering to a question which seems to be simple yet more complex while approaching "can a computer mimic the human nervous system?". Human brain is highly complex nonlinear parallel computer. Structural constituents of human brain are known as neurons which are interconnected in complex way (Lee, T.S.,, Mumford, D.2003). The task of recognition through a neural network achieved by teaching the computer what to do in each and every step with a set of pre-defined procedures that is it is more like if a computer has to identify a photograph first it has to be taught by feeding various photographs then it go for a matching with the original photograph with the ones which it stores in the large database repository. It gives out the picture which carries a best resemblance with original one. While answering to the question another constraint researchers faced is human brain processing speed which is of 10^{-3}s where as a silicon IC response time is 10^{-9}s.

The evolution of the deep machine learning starts from convolutional neural networks, deep neural networks and deep belief networks eras. Convolutional neural networks are a kind of special multi- layer neural networks which are inspired from biology. It is similar to other neural networks and they are trained with a version of back propagation algorithm but only differ in the architecture. CNN are mainly designed to recognize visual patterns directly from pixel images with minimal preprocessing. The hurdles faced by feed forward neural networks towards image processing could be resolved by the entry of CNN to some extent. The major constraint that normal FFNN facing was the rapidly growing weights with the dimensionality of the input, since every object in one unit is connected to every object in next unit. That makes it slow in learning for high dimensional domain of vision. Learning to recognize an object wouldn't even propagate the same object presented in different visual field, it's because the every pair of neurons between different units has separate weights which would be involved in the calculation. A CNN comprises of various convolutional and sub sampling layers alternatively took after by completely joined layers. The input to a convolutional layer is an m x m x r picture where m is the tallness and width of the picture and r is the quantity of channels, e.g. a RGB picture has r=3. The convolutional layer will have k layers (or portions) of size n x n x q where n is littler than the measurement of the picture and q can either be the same as the quantity of layers or littler and may shift for every part. The measure of the channels offers ascend to the generally joined structure which are each convolved with the picture to create k highlight maps of size m−n+1. Every guide is then sub sampled ordinarily with mean or max pooling over p x p adjoining districts where p ranges between 2 for little pictures (e.g. MNIST) and is normally not more than 5 for bigger inputs. Either before or