International Research Journal of Engineering and Technology (IRJET) e-ISSN:2395-0056
Volume: 04 Issue: 11 | Nov -2017 www.irjet.net p-ISSN:2395-0072

# Deep Learning for Natural Language Processing

**Sachee Nene[1]**

*[1]Student,Dept. of Computer Engineering, VESIT, Maharashtra, India*

---

**Abstract -** *Deep Learning has come into existence as a new area for research in Machine Learning. It aims to act like a human brain, having the ability to learn and process from complex data and also tries solving intricate tasks as well. Due to this capability, its been used in various fields like text, sound, images etc. Natural language process has started to being impacted by the deep learning techniques. This research paper highlights Deep Learning's recent developments and applications in Natural Language Processing.*

***Key Words:*** **Deep Neural Network (DNN) , Support Vector Machines(SVM), Restricted Boltzmann Machine(RBM),Conditional Random Field (CRF), Natural Language Processing(NLP),**

## 1.INTRODUCTION

From 2006 , deep learning has come up as a new research area of machine learning. Deep learning can be defined as a set of machine learning algorithms that try to learn inputs from multiple layered models , such as neural networks. DNNs comprises of many layers of non linear operations. Before the emergence of deep learning in 2006,searching deep architectures parameter space was a significant task, but with the latest deep learning algorithms it is easier to solve this problem with high value of success.

## 2.DEEP LEARNING

The main idea of Deep Learning is referred to as a process to learn a structure of features a level at a time, and this is called as greedy layer wise unsupervised pre training. This process can be entirely unsupervised and tries to learn from the features of previous levels, to obtain and rebuild the original data. The pre training underlying principle is to train each layer with an unsupervised algorithm ,by taking the output of the previous level as an input to the next level. This can be used either as an input to SVM or CRF or it can be used as an initialization for a deep neural network. Both of them are in supervised form. The weights are added to a Deep Neural Network by each layer of unsupervised learning process. In sum ,the stack of learned weights can be used to initialize a neural network classifier which is a predictor or a deep generative model such as Deep Boltzmann Machine.

### 2.1 Stacked auto-encoder

A good example of greedy layer wise unsupervised pre training is a stack auto-encoder. It is a neural network which consists of multiple layers of sparse auto encoders in which outputs of previous layers is given as a input to every next layer. This encoder is specifically used to learn a complex representation of a set data which is extensively used for data compression and dimensionality reduction. Initially, consider a stacked encoder with n layers. The general layers involve : an input layer, a set of small hidden layers(for encoding),and an output layer for reconstructing the input layer. Let $W^{(k,1)},W^{(k,2)},b^{(k,1)},b^{(k,2)}$ denote the parameters $W^{(1)},W^{(2)},b^{(1)},b^{(2)}$ for kth encoder. The most important step in this model is the Training phase. In this, the first layer is trained to get parameters $W^{(1,1)},W^{(1,2)},b^{(1,1)},b^{(1,2)}$.This first layer converts the raw input into a vector of activation function of hidden layers. Similarly ,the second layer is trained to get parameters $W^{(2,1)},W^{(2,2)},b^{(2,1)},b^{(2,2)}$.In this way the process is continued up to kth layer accepting the parameters $W^{(k,1)},W^{(k,2)},b^{(k,1)},b^{(k,2)}$.

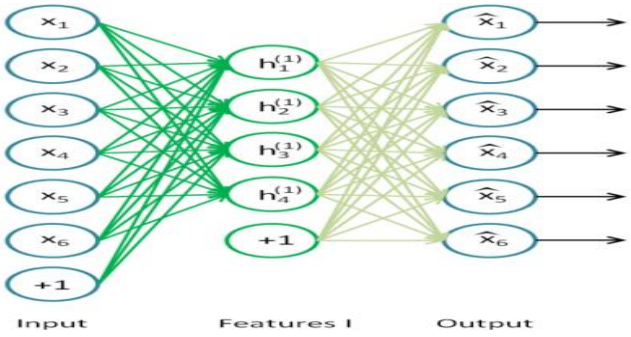


**Fig -1**: Structure of Auto-encoder

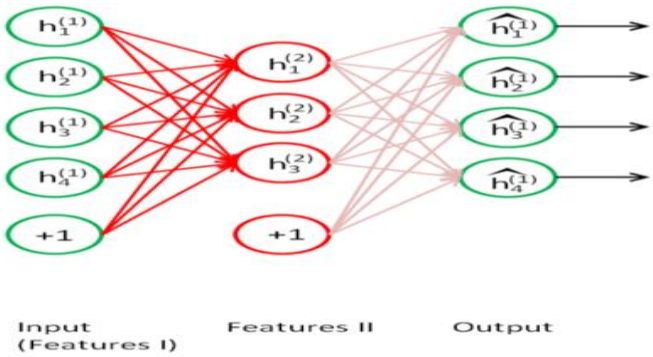


**Fig -2**: Stacked auto-encoder

In between the layers of the model there is an encoding step in forward order.It is given as :

$$a^{(1)} = f(z^{(l)})$$
$$z^{(l+1)} = W^{(l,1)}\, a^{(l)} + b^{(l,1)}$$

The decoding step in reverse order is given as :

$$a^{(n+1)} = f(z^{(l)})$$
$$z^{(n+l+1)} = W^{(n-l,2)}\, a^{(n+l)} + b^{(n-l,2)}$$

The actual information is present in $a^{(n)}$, which is the activation of hidden inputs.

In conclusion, this method trains parameters of individual layers and freezes the remaining parameters of the model. To produce more fine tuned result , fine-tuning using back propagation is used at the end of training phase. It improves results by tuning parameters of all layers at the same time.

## 2.2 Deep Boltzmann Machines

Additional method to implement the pre training is to use the RBM. It is used for classification, collaborative filtering, dimensionality reduction, regression. These two layer neural nets build up the deep belief networks. Layer 1 is the visible/input layer ,and layer 2 is the hidden layer.

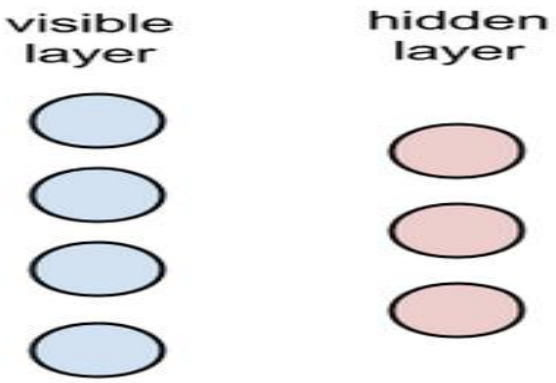


**Fig -3**: Structure of Restricted Boltzmann Machine

The circle represents a node and calculations are present in it. These nodes are connected via layers, but there is no intra layer communication. Every node takes a low level feature for learning from an item in the dataset. Inputs to the next layer nodes are given as learned feature activations. Once the pre training is over, RBM is unveiled for the creation of deep networks. By using the error back propagation algorithm it is fine tuned. Deep Boltzmann Machines are formed by the stacks of RBM. Initialization of Deep Boltzmann Machines is done by pre trained DBM and trained with stack auto encoder.

Example of Restricted Boltzmann Machine :
Consider a dataset of gray scale images. Each node will obtain one pixel value for each pixel in an image. The network is a two layer network. Node 1 of hidden layer is multiplied by a weight and added to bias. The obtained result is given to an activation function which produces the output.

Activation f((weight w* input x)+bias) = output a

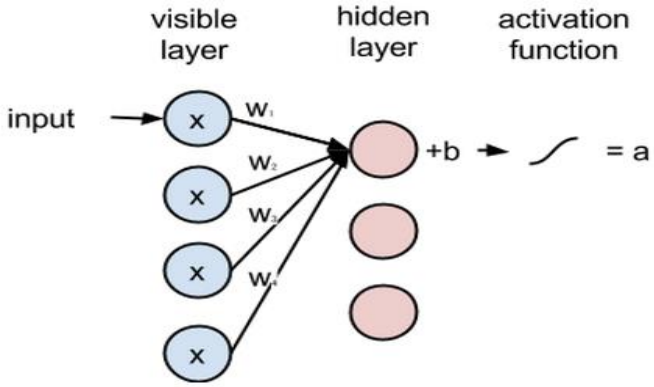


**Fig -4**: Weighted inputs combined at hidden nodes

Each node is multiplied by a weight. products are added together, and this is summed with bias and this result is sent through an activation function to produce the output.

So, each input will have three weights and in total for the network there will be 12 weights. Matrix is formed for the weights between two layers in which rows represent the inputs and columns represent the outputs. The hidden nodes receive the four inputs with multiplied weights. This sum is again added with the bias and passed through the activation algorithm to become the output of the hidden node

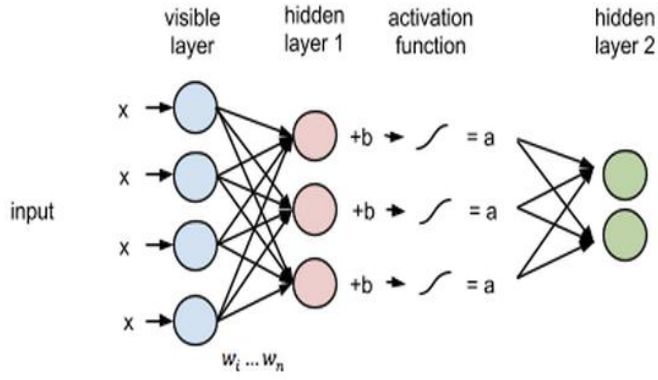


**Fig -5**: Multiple hidden layers

### 2.3 Why deep?

Using the real context meaning of deep that is extending far down from the top or surface helps to represent a non linear function with fewer parameters. Due to the involvement of large complexity results, functions can be compacted using deeper architecture .As each parameter of the architecture might have to be learned or used, deeper architecture can be more beneficial to obtain high statistical efficiency. Deeper architectures also permit hierarchical representations.

Deep learning is inspired from the architecture of brain. Brain simulations are used to make algorithms better and easier. With the availability of more advance and fast computers it is possible to design large neural networks and train them with myriad amounts of data. There can also be seen a high increase in the performance.

### 2.4 Multi-Task and Transfer Learning, Domain Adaptation

Multi-Task learning involves multiple learning tasks which are solved at the same time, utilizing the differences and commonalities across different tasks. Improved learning efficiency and prediction accuracy can be obtained. By jointly learning the performance of multiple classification tasks it improves the performance.

## 3.APPLICATIONS OF DEEP LEARNING

Deep learning has already left a mark on machine learning and artificial intelligence .It is brought along with machine learning to achieve artificial intelligence. It is started being used in fields such as sounds, pictures, motion and text. The flourishment in areas of academics and industry has led to the increase in scientific research on deep learning.

### 3.1 Object Recognition

Object recognition is a nontrivial task for a computer and this is why deep learning has moved from digits recognition to object recognition in pictures. Object recognition deals with identification of a particular object in an image or a video. Its algorithm makes use of learning, matching, pattern recognizing algorithms to know about feature or appearance.

There are two approaches involved. First, involves appearance based method which uses example images of the objects to perform recognition. Objects look different under conditions like color, viewing direction, sizing or shape. The different ways of achieving them are edge matching, grey scale matching, gradient matching.

Second, is the feature based method that involves a search to find possible matches between object features and image features. The features used for comparison are surface patches, corners and linear edges. The different ways include hypothesize and test, invariance, geometric hashing.

Object recognition is being used in a variety of applications like disease identification, systems for driver assistance ,face detection, image watermarking, visual positioning & tracking and few are also based on through the use of neural networks and linear binary patterns.

### 3.2 Speech Recognition and Signal Processing

Speech recognition is the process of identifying words and phrases in any spoken language which is done by the machine and convert it into machine readable form. It can be regarded as an alternative to usual methods of interacting with the computer that is giving an input through the keyboard. Hence, by using speech recognition system it will reduce or replace the dependency on standard input.

The current system of speech recognition uses complex components. The current system using deep learning uses certain learning algorithms which learns features without assuming. It consists of four basic attributes : processing units which consists of input and computational unit, connections which include hidden layers with 100 units each in which 93 and input and 43 are output units, computation involves producing an input pattern to the network ,training and training procedure to make the connections adaptable so the network can produce the required output pattern from the input.

There are several advantages that make neural network a better system for speech recognition. The advantages of neural networks are they are powerful,self adjusting and its recognition pattern is well sophisticated. The most crucial requirement is the data and it is available in large amounts with voice data.

### 3.3 Machine Translation

Machine translation is the process that deals with translation of speech or text from one language to another. It involves substitution of words in one language into the required one which is done with the help of corpus statistical and neural techniques that leads to better translation which clearly handles differences in language.

The current translation techniques involve certain limitations but with neural networks it produces improved quality output as there is some human intervention for recognition of words unambiguously. The process is simple

like decoding the meaning of source text and re-encoding the meaning in the target language. The different approaches are rule based such as dictionary based, statistical such as example based and hybrid machine translation which is a combination of rule based and example based.

Machine translation is used in everyday applications such as Skype, Google talk, MSN Messenger.

### 3.4 Question and Answering

Question and Answering is the process of building systems that automatically answer's questions presented by humans. It is also used for entity extraction and retrieving of information. It is used in chat bots and systems involving human conversation. The main element of this application is the question and answering dataset. There are two types :open and closed. In open datasets, the answer is produced based on general knowledge and information provided along with the dataset if any. The Quiz Bowl and Google N-Grams are examples of open datasets. In closed datasets, the answer is produced based on the information provided in the dataset itself.

In earlier systems a pipelined structure was used which comprised of NLP techniques like co-reference resolution, parsing, part-of-speech tagging and it was very complex.

In the recent systems involving deep learning and neural networks the pipelined structure involved is more simple but requires a good amount of training. This allows for handling longer question length, improved memory and greater relevance with general answering required for answering. The current systems are more promising than the previous in terms of relevant and complete answering to a particular question.

IBM's Watson is most eminent real world application of question and answering.

## 4.CONCLUSIONS

Deep learning is the most emerging field of machine learning and has shown extensive contribution in various research areas. It has helped in overcoming the drawbacks of traditional methods by making the systems less complex and fast. Since deep has a solid practical background, strong theory based fundamentals also need to be established. Deep learning has been used with NLP in several research areas, which is quite promising and is a success. Though the outcomes are still in the rudimentary stages with some research areas ,the systems understand the human language more better. This requires a more massive research on both deep learning and NLP.

## REFERENCES

[1] Josh Patterson and Adam Gibson,Deep Learning : A Practitioner's Approach,O'Reilly Media,CA,2017

[2] Laurene Fausett,Fundamentals of Neural Networks,Pearson,2004

[3] Itamar Arel, Derek C. Rose, and Thomas P. Karnowski, "Deep Machine Learning – A New Frontier in Artificial Intelligence Research",MCI,Nov. 2010, pp. 1556-603X,doi:10.1109/MCI.2010.938364

[4] Juergen Schmidhuber, "Deep Learning in Neural Networks: An Overview",Neural and Evolutionary Computing,vol.61, Jan. 2015, pp. 85-117,doi:10.1016/j.neunet.2014.09.003

[5] Yann LeCun,Yoshua Bengio and Geoffrey Hinton, "Deep Learning",Nature,vol.521,May. 15,pp. 436-444,doi:10.1038/nature14539

[6] Ruslan Salakhutdinov and Geoffrey Hinton, "An Efficient Learning Procedure for Deep Boltzmann Machines"NeuralComputation,vol. 24,2012,pp. 1967-2006

[7] Dong Yu and Li Deng, Automatic Speech Recognition - A Deep Learning Approach,Springer,London2015

[8] Minwei Feng,Bing Xiang and Bowen Zhou, "Distributed Deep Learning for Question Answering",vol.3,Aug.16,1511.01158

[9] Christian Szegedy Alexander Toshev Dumitru Erhan, "Deep Neural Networks for Object Detection"

[10] W.John Hutchins and Harold L. Somers,An introduction to machine translation,Academic Press,London,1992