ISSN 2277 – 5528
Impact Factor- 4.015

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & MANAGEMENT

A STUDY & REVIEW ON TECHNIQUES, APPLICATIONS AND CHALLENGES USED IN DEEP LEARNING

Khushboo Mandhaniya[*1] & Bhushit Chandra Nema[2]
[*1]Electronics & Telecommunication, Institute of Engineering & Technology (DAVV), Indore (M.P.), India
[2]Electronics & Communication, Ujjain Engineering College, Ujjain (M.P.), India

ABSTRACT

This paper study and survey the use of deep learning techniques, applications and challenges that we are facing in today's era. The technologies are upgrading very rapidly in the field of higher education. Now we are in the world of big data. In this world of big data, we have variety, velocity and volume of data to be available. To handle such a huge data in efficient manner is a complex task for IT organizations as well as researchers. Deep learning is an emerging research area in machine learning and pattern recognition field. Deep learning refers to machine learning techniques that use supervised or unsupervised strategies to automatically learn hierarchical representations in deep architectures for classification. The objective is to discover more abstract features in the higher levels of the representation, by using neural networks which easily separates the various explanatory factors in the data. In the recent years it has attracted much attention due to its state-of-the-art performance in diverse areas like object perception, speech recognition, computer vision, collaborative filtering and natural language processing.

Keywords: *Deep Learning, Applications, Techniques, Challenges*

I. INTRODUCTION

Machine learning has successfully grown to be a major computer science discipline with extensive applications in science and engineering for many years. The computer extracts knowledge through supervised experience, where a human operator is involved in helping the machine learn by giving it hundreds or thousands of training examples, and manually correcting its mistakes. While machine learning has become leading within the field of AI, it does have its problems. It is particularly time consuming and is still not a true measure of machine intelligence as it relies on human resourcefulness to come up with the abstraction that allow computer to learn. A primary challenge to machine learning is the lack of adequate training data to build accurate and reliable models in many realistic situations. When quality data are in short supply, the resulting models can perform very poorly on a new domain, even if the learning algorithms are best chosen. Unlabeled data is cheap and plentiful, unlike labeled data which is expensive to obtain. The promise of self-taught learning is that by exploiting the massive amount of unlabeled data, much better models can be learnt. By using unlabeled data to learn a good initial value for the weights in all the layers, the algorithm is able to learn and discover patterns from massive amounts of data than purely supervised approaches. This frequently results in much better classifiers being learned.

Deep learning is mostly unsupervised contrasting machine learning which is supervised. It involves creating large scale neural nets that allow the computer to learn and compute by itself without the need for direct human intervention. Learning in machine learning applications depends on hand-engineering features where the researcher manually encodes relevant information about the task at hand and then there is learning on top of that. This contrasts with deep learning which tries and gets the system to engineer its own features as much as is viable. The recent Google experiments on deep learning have shown that it is possible to train a very large unsupervised neural network to automatically develop features for recognizing cat faces. The data scarcity problem associated with extremely large-scale recommendation systems provides strong motivation for finding new ways to transfer knowledge from auxiliary data sources. Deep Learning is a new area of Machine Learning research, which has been introduced with the objective of moving Machine Learning closer to one of its original goals: Artificial Intelligence. Deep Learning is about learning multiple levels of representation and abstraction that help to make sense of data such as images, sound, and text.

In this paper, we will discussing the terminologies used in Deep learning in Section II, Literature Survey in Section III, Conclusion and Future Enhancement in section IV and finally we have used the references.

II. BASIC TERMINOLOGIES USED IN DEEP LEARNING

1. **Deep belief network (DBN):** Probabilistic generative models composed of multiple layers of stochastic, hidden variables. The top two layers have undirected, symmetric connections between them. The lower layers receive top-down, directed connections from the layer above [1, 2].
2. **Boltzmann machine (BM):** A network of symmetrically connected, neuron-like units that make stochastic decisions about whether to be on or off [1, 2].
3. **Restricted Boltzmann machine (RBM):** A special BM consisting of a layer of visible units and a layer of hidden units with no visible-visible or hidden-hidden connections [1, 2].
4. **Deep Boltzmann machine (DBM):** A special BM where the hidden units are organized in a deep layered manner, only adjacent layers are connected, and there are no visible-visible or hidden-hidden connections within the same layer [1, 2].
5. **Deep neural network (DNN):** A multilayer network with many hidden layers, whose weights are fully connected and are often initialized (pre-trained) using stacked RBMs or DBN [1, 2].
6. **Deep auto-encoder:** A DNN whose output target is the data input itself, often pre-trained with DBN or using distorted training data to regularize the learning [1, 2].
7. **Distributed representation:** A representation of the observed data in such a way that they are modeled as being generated by the interactions of many hidden factors. A particular factor learned from configurations of other factors can often generalize well. Distributed representations form the basis of deep learning [1, 2].

III. LITERATURE SURVEY

There were attempts at training deep architectures before 2006 but failed because training a deep supervised feed forward neural network yielded worse results both in training and in test error than shallow ones with 1 or 2 hidden layers. The scenario was changed by three important papers by Hinton, Bengio and Ranzato[3],[4],[5]. The key principles found in all three papers are on unsupervised learning of representations used to pre-train each layer. The unsupervised training in these works is done one layer at a time, on top of the previously trained ones. The representation learned at each level is the input for the next layer. Then supervised training is used to fine-tune all the layers.

Geoffrey Hinton [3] trained deep belief networks by stacking Restricted Boltzman Machines (RBMs) on top of one another as deep belief network. The Deep Belief Networks use RBMs for unsupervised learning of representation at each layer.

The Bengio [4] paper explores and compares RBMs and auto-encoders. The Ranzato [5] et al paper uses sparse auto-encoder in the context of a convolutional architecture. Recently notable progresses have been made to lessen the challenges related to high data volumes. When there is huge volume of data it is often impossible to train a deep learning algorithm with a central processor and storage. Hence distributed frameworks with parallelized machines are ideal.

Deng et al. [6] proposed a modified deep architecture called Deep Stacking Network (DSN), which can be parallelized. A DSN is a combination of several specialized neural networks with a single hidden layer. Stacked modules with inputs composed of raw data vector and the outputs from previous module form a DSN.A new deep architecture called Tensor Deep Stacking Network (T-DSN), which is based on the DSN, is implemented using CPU clusters for scalable parallel computing. Recent models make use of clusters of CPUs or GPUs to increase the training speed. Deep learning algorithms possess one of the unique characteristics of using unlabeled data during training. Training with vastly more data is preferable to using smaller number of exact, clean, and carefully curated data, though incompleteness and noisy labels are part of data. To address the the effect of noisy labels, a more efficient cost function and novel training strategy may be needed.

Gravier, Garg[7,11] : survey presents Visual speech information from the speaker"s moth region has been successfully shown to improve noise robustness of automatic speech recognizers , thus promising to extend their usability into the human computer interface. In this paper, we review the main components of audio-visual automatic speech recognition and present novel contributions in two main areas: first, the visual front end design and later, we discuss new work on features and design fusion combination , the modeling of audio-visual speech asynchrony and incorporating modality reliability estimates to the bimodal recognition process.

Das[8]: presents a brief survey on speech is the primary and the most convenient means of communication between people. The communication among human computer interaction is called human computer interface. Speech has potential of being important mode of interaction with computer. This paper gives an overview of major technological perspective and appreciation of the fundamental progress of speech recognition and also gives overview technique developed in each stage of speech recognition. This paper helps in choosing the technique along with their relative merits and demerits. A comparative study of different techniques is done. This paper concludes with the decision on feature direction for developing technique in human computer interface system in different mother tongue and it also gives the various technique used in each step of a speech recognition process and attempts to analyze an approach for designing an efficient system for speech recognition . The objective of this review paper is to summarize and compare different speech recognition systems and identify research topics and applications where are at the front end of this exciting and challenging field.

Dhameliya, Desai[9,10]: survey presents speech is the most natural form of human communication and speech processing has been one of the most inspiring expanses of signal processing . Speech recognition is the process of automatically recognizing the spoken words of person based on information in speech signal. Automatic Speech Recognition(ASR) system takes a human speech utterances as an input and requires a string of words as output. This paper introduces a brief survey on Automatic Speech Recognition and discusses the major subjects and improvements made in the past 60 years of research , that provides technological outlook and a respect of fundamental achievements that have been accomplished in the important areas of speech recognition. Definition of various types of speech classes , feature extraction techniques, speech classifiers and performance evaluation are issues that require attention in designing of speech recognition system. The objective of this review paper is to summarize some of the well-known methods used in several stages of speech recognition system.

IV. TECHNIQUES & APPLICATIONS OF DEEP LEARNING

RBM: Boltzmann Machines (BMs) are a particular form of log-linear Markov Random Field (MRF), i.e., for which the energy function is linear in its free parameters. To make them powerful enough to represent complicated distributions (i.e., go from the limited parametric setting to a non-parametric one), we consider that some of the variables are never observed (they are called hidden). By having more hidden variables (also called hidden units), we can increase the modeling capacity of the Boltzmann Machine (BM). Restricted Boltzmann Machines further restrict BMs to those without visible-visible and hidden-hidden connections. A graphical depiction of an RBM is shown below.

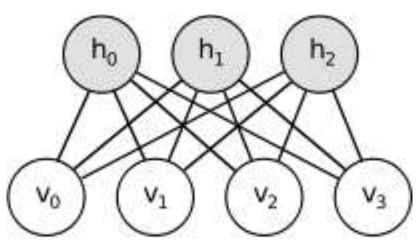

Figure 1: Architecture of Boltzmann Machine

The energy function $E(v,h)$ of an RBM is defined as:

$$E(v,h) = -b'v - c'h - h'Wv$$

where W represents the weights connecting hidden and visible units and b, c are the offsets of the visible and hidden layers respectively.

This translates directly to the following free energy formula:

ISSN 2277 – 5528
Impact Factor- 4.015

$$\mathcal{F}(v) = -b'v - \sum_i \log \sum_{h_i} e^{h_i(c_i + W_i v)}.$$

Because of the specific structure of RBMs, visible and hidden units are conditionally independent given one-another. Using this property, we can write:

$$p(h|v) = \prod_i p(h_i|v)$$
$$p(v|h) = \prod_j p(v_j|h).$$

Convolution Neural Network: Convolution neural networks. Sounds like a weird combination of biology and math with a little CS sprinkled in, but these networks have been some of the most influential innovations in the field of computer vision. 2012 was the first year that neural nets grew to prominence as Alex Krizhevsky used them to win that year's Image Net competition (basically, the annual Olympics of computer vision), dropping the classification error record from 26% to 15%, an astounding improvement at the time. Ever since then, a host of companies have been using deep learning at the core of their services. Facebook uses neural nets for their automatic tagging algorithms, Google for their photo search, Amazon for their product recommendations, Pinterest for their home feed personalization, and Instagram for their search infrastructure.

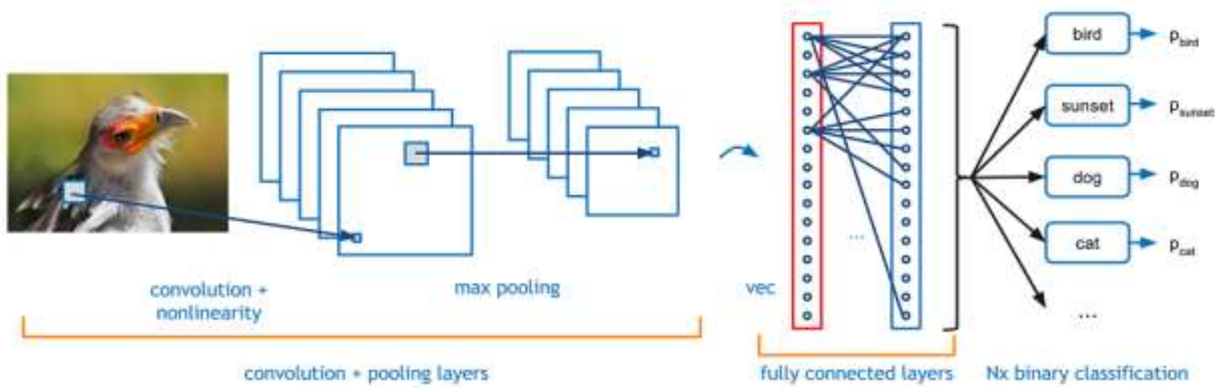

Figure 1: Example of Convolution Neural Network

Deep Recursive Neural Network: Recursive neural networks comprise a class of architecture that can operate on structured input. They have been previously successfully applied to model compositionality in natural language using parse-tree-based structural representations. Even though these architectures are deep in structure, they lack the capacity for hierarchical representation that exists in conventional deep feed-forward networks as well as in recently investigated deep recurrent neural networks. In this work we introduce a new architecture — a deep recursive neural network (deep RNN) — constructed by stacking multiple recursive layers. We evaluate the proposed model on the task of fine-grained sentiment classification. Our results show that deep RNNs outperform associated shallow counterparts that employ the same number of parameters. Furthermore, our approach outperforms previous baselines on the sentiment analysis task, including a multiplicative RNN variant as well as the recently introduced paragraph vectors, achieving new state-of-the-art results. We provide exploratory analyses of the effect of multiple layers and show that they capture different aspects of compositionality in language.

Auto encoder: An auto encoder is classically a feed forward neural network which aims to learn a compressed, distributed representation of a dataset. An auto-encoder is a 3 layer neural network, which is trained to reconstruct its inputs by using them as the output. It needs to learn features that capture the variance in the data so it can be reproduced. It can be shown to be equivalent to PCA, if linear activation functions are only used and can be used for dimensionality reduction. Once trained, the hidden layer activations are used as the learned features, and the top layer can be discarded. Auto encoders are trained using the strategies like de-noising, contraction and sparseness.

During de-noising in Auto-Encoders some random noise is added to the input. The encoder is required to reproduce the original input. Randomly deactivating inputs during training will improve the generalization performance of regular neural networks. In contractive Auto-Encoders, setting the number of nodes in the hidden layer to be much lower than the number of input nodes forces the network to perform dimensionality reduction. This prevents it from learning the identity function as the hidden layer has insufficient nodes to simply store the input. Sparse Auto-Encoders are trained by applying a sparsity penalty to the weight update function. It penalizes the total size of the connection weights and causes most weights to have small values. RBM's or Auto-Encoders can be trained layer by layer. The features learned from one layer are fed into the next layer, so that first a network with 1 hidden layer is trained, and only after that is done, a network with 2 hidden layers is trained, and so on. At each step, the old network with k- 1 hidden layers is taken and an additional k-th hidden layer is added that takes as input the previous hidden layer $k - 1$ that was trained.

V. CONCLUSION

The results of this study suggest that these items, when combined with existing core survey items, assess three distinct aspects of a second order factor that, in content, appears to be related to deep learning. Natural Language Processing (NLP) is a typical example; deep learning cannot understand a story, as well as a general request to an expert system. So there's still a long way to go before we can implement the real intelligent machine. But deep learning indeed provides a direction to implement the more intellectual learning; therefore it can be regarded as a small step toward AI. Deep architectures help deep learning by trading a more complicated space for better performance, in some cases, even for less computation time. Deep architectures are good models for deep learning, but can't be proved to be the best one. There're still many possibilities in the architectures and learning algorithms that can carry out better performances

REFERENCES

[1] Lei zheng. 2016. A survey and critique of deep learning on recommender systems. Department of computer science. Li Deng, Dong Yu. 2014. Deep Learning: Methods and Applications. Foundations and Trends in Signal Processing.
[2] Li Deng, Dong Yu. 2014. Deep Learning: Methods and Applications. Foundations and Trends in Signal Processing.
[3] Hinton, G. E., Osindero, S. and Teh, Y., " A fast learning algorithm for deep belief nets", Neural Computation 18:1527-1554, 2006
[4] Yoshua Bengio, Pascal Lamblin, Dan Popovici and Hugo Larochelle, "Greedy Layer-Wise Training of Deep Networks", in J. Platt et al. Advances in Neural Information Processing Systems 19 (NIPS 2006), pp. 153-160, MIT Press, 2007
[5] Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopra and Yann LeCun "Efficient Learning of Sparse Representations with an Energy-Based Model", in J. Platt et al. (Eds), Advances in Neural Information Processing Systems (NIPS 2006), MIT Press, 2007
[6] Li Deng, Xiaodong He, Jianfeng Gao, Deep stacking networks for information retrieval , Acoustics, speech and signal processing, 2013 IEEE International conference.
[7] A survey on deep learning: one small step toward AI 2012. Xavier Glorot, Antoine Bordes et. al : Domain Adaptation for Large-Scale Sentiment Classi_cation:A Deep Learning Approach, 2008.
[8] ALISON RUSHTON, A. University of Birmingham, Edgbaston,Birmingham, UK : Formative assessment: a key to deep learning? , 2007.
[9] Dean, J. and Ghemawat, S. Map-Reduce: simplified data processing on large clusters. Comm. ACM, pp. 107–113,2008.
[10]Deng , Li,Gong,et al. : Proximal regularization for online and batch learning. In ICML, 2009.
[11]Gaikwad, Gawali and Yannawar : A practical guide to training restricted boltzmann machines. Technical report, U. of Toronto, 2010