International Journal of Enhanced Research in Science Technology & Engineering, ISSN: 2319-7463
Vol. 4 Issue 5, May-2015, pp: (178-182), Impact Factor: 1.252, Available online at: www.erpublications.com

# Deep Learning - A Way to Artificial Intelligence

Shivani Khurana[1], Chhavi Rana[2]
[1]UIET MDU, Rohtak, Haryana
[2]Asstt. Prof., UIET MDU, Rohtak, Haryana

**Abstract: Deep learning is one of the emerging fields belonging to Artificial Intelligence. Deep learning is combination of neural network, optimization, graphical modeling, pattern recognition, and signal processing. The goal of deep learning is to build a model which is capable to process complex input data, learn different features fast and intellectually, and can solve efficiently different kinds of complex tasks by mimicking the capabilities of human brain. The proposed techniques in this area should deal with high dimensionality of data which is a hurdle in many science and engineering applications. This paper presents a survey on different deep architectures and their application areas.**

**Keywords: Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Deep Belief Network (DBN), Deep Learning, Machine Learning.**

## I. Introduction

Artificial Intelligence deals with building a computer which can sense, remember, learn, and recognize things like a human being. The first machine which can sense and learn is perceptron, but it has limited learning abilities. Neural network with multiple hidden layers can easily learn more complicated functions, but learning is a very slow process. Support vector machines (SVM) can easily solve many practical AI problems, but it is not a good choice due to its shallow architecture [1]. A standard neural network consists of an input layer, a number of hidden layers, and an output layer. Input neurons get activated through sensory inputs from environment and weighted connections from previously active neurons are used to activate other neurons. Deep learning deals with accurate credit assignment that is finding the weights which are used to activate such other neurons so that neural network can exhibit desired behavior [2].

Deep learning is a class of machine learning techniques [3] with the deep architecture and the good learning algorithms, which can perform different intellectual tasks [1]. Deep learning is combination of neural network, optimization, graphical modeling, pattern recognition, and signal processing [3]. The various application areas of deep learning are computer vision, voice search, phonetic recognition, hand-writing recognition, conversational speech recognition, speech and image feature coding, audio processing, semantic utterance classification, visual object recognition, and information retrieval [3]. The proposed techniques in this area should deal with high dimensionality of data which is a hurdle in many science and engineering applications. The dimensionality can be reduced by pre-processing the data and this scheme is known as feature extraction [4].

The remainder of this survey is organized in the following way: Section II and section III gives a brief introduction to Convolutional Neural Network (CNN) and Deep Belief Network (DBN) respectively. Section IV includes some of the applications of deep learning. Section V includes challenges in the field of deep learning and in section VI we provide the overall conclusion of this survey and future scope of deep learning.

## II. Convolutional Neural Network

CNN belongs to the multi-layer feed-forward neural network particularly designed for use on two-dimensional data, such as videos and images. CNN is influenced by time-delay neural networks (TDNN) where weights are shared in temporal dimension which reduces learning computation requirements [5] and it is the first successful deep learning approach where multiple layers are trained in an efficient manner [4]. CNN combines local receptive field, shared weights, and spatial or temporal subsampling to ensure some degree of shift and distortion invariance. Local receptive field or shift window is small portion of the data and is treated as inputs to the lowest layer of the hierarchical structure. Digital filtering is applied at each layer in order to obtain important features of the data observed. Shared weights reduces the number of free parameters being learnt which increases the learning efficiency. Dimensionality of feature maps is reduced by temporal or spatial subsampling, and now the convolution process is applied to the new set of feature maps [1]. In fig. 1, sparse connectivity in CNN is illustrated. Layer m-1 is the input layer. The neurons in layer m have receptive field of width 3 with respect to input layer because each neuron is connected to three adjacent neurons in input layer. Similarly neurons in layer m+1 have receptive field of width 3 with respect to layer m, but receptive field with respect to layer m-1 is 5.

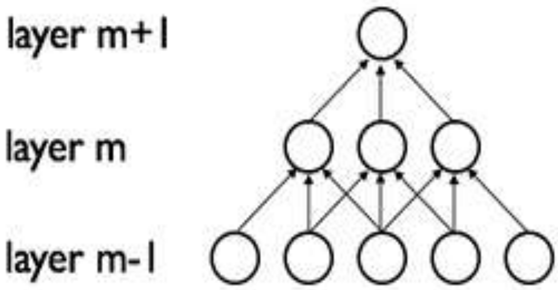


**Figure 1. Sparse connectivity [1], [6]**

Sharing of weights is shown in fig. 2. The weights shown in the same color are shared and are identical. Learning efficiency is increased by sharing the weights and thus reducing the number of free parameters being learnt.

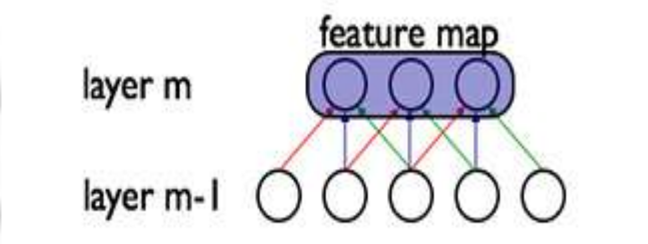


**Figure 2. Sharing of weights in CNNs [1], [6]**

The CNN consists of convolution process and subsampling process. Fig. 3 shows convolution and subsampling process. In convolution process, an input is convolved with a trainable filter fx and then a trainable bias bx is added to produce the convolution layer Cx. In subsampling process, summing of neighborhood (four pixels) is done which is followed by weighting by scalar wx+1. After that trainable bias bx+1 is added and the resultant is passed through sigmoid function which produces a smaller feature map Sx+1.

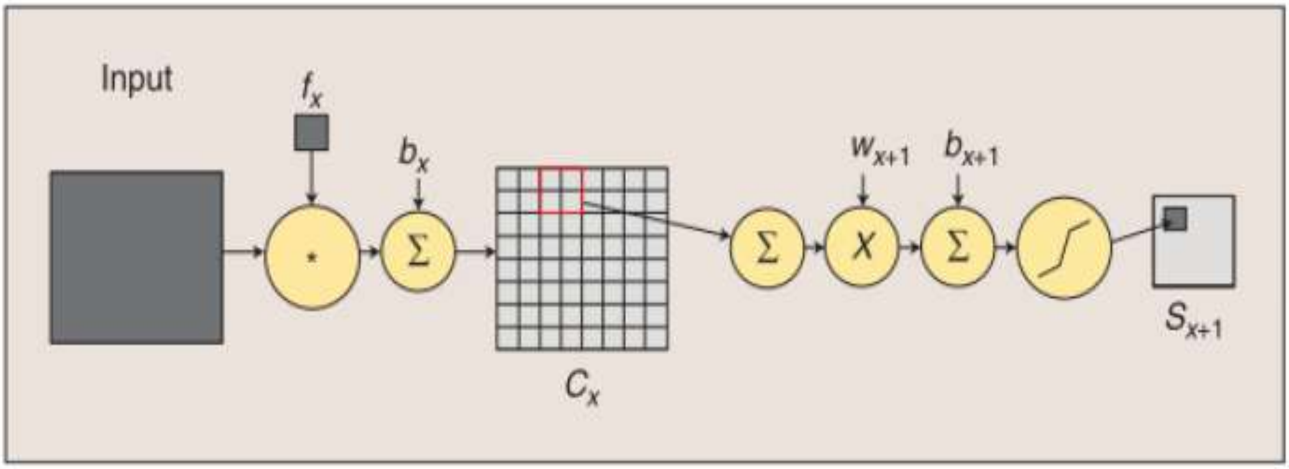


**Figure 3. The convolution and subsampling process [4]**

Fig. 4 shows a conceptual example of convolutional neural network. A sequence of convolution and subsampling process is applied on input to produce final output "NN". The dimensions of feature maps are decreased after each layer. Smaller value of weight can cause blurring of the image, output can also resemble an AND or OR function [4]. Back-propagation is used as learning algorithm to update the weights and one complete cycle as shown in fig. 4 result in weight updation [1]. In CNN, invariance to object translations is created by a method known as feature pooling [7] and it is handcrafted by the network organizer. Final outputs are produced by forwarding the activation outputs to conventional feedforward neural network [4]. Recent research areas of CNN are speech detection [8], document analysis [9], face detection [10] [11], license plate detection [11], and temporal coherence in videos [12].

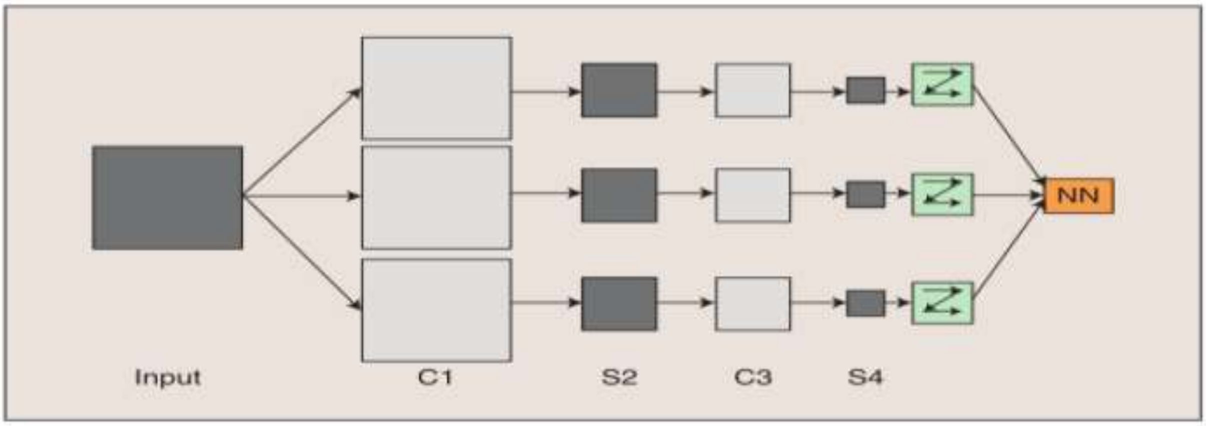


**Figure 4. Complete working of conceptual neural network [4]**

## III. Deep Belief Network

Fig. 5 shows the structure of DBN which consist of two parts that is associative memory and Restricted Boltzman Machine (RBM) layers. RBM is directed graph model [1] consisting of single visible layer connected to a single hidden layer where neurons within a layer are not connected [4]. The top two layers form the associative memory which is undirected graph model [1]. The two variations of DBN are generative model and discriminative model [1]. Generative model facilitates the estimation of both P (Observation|Label) and P (Label| Observation), and it is a top-down procedure represented by green arrow in the fig. 5. Discriminative model provides estimation of only P (Label|Observation), and it is bottom-up procedure represented by red arrow in the fig. 5 [1] [4]. "Layer-by-layer" learning strategy is applied in DBN where previous layers are used to learn features of higher-level layers and these features are more complicated [1].

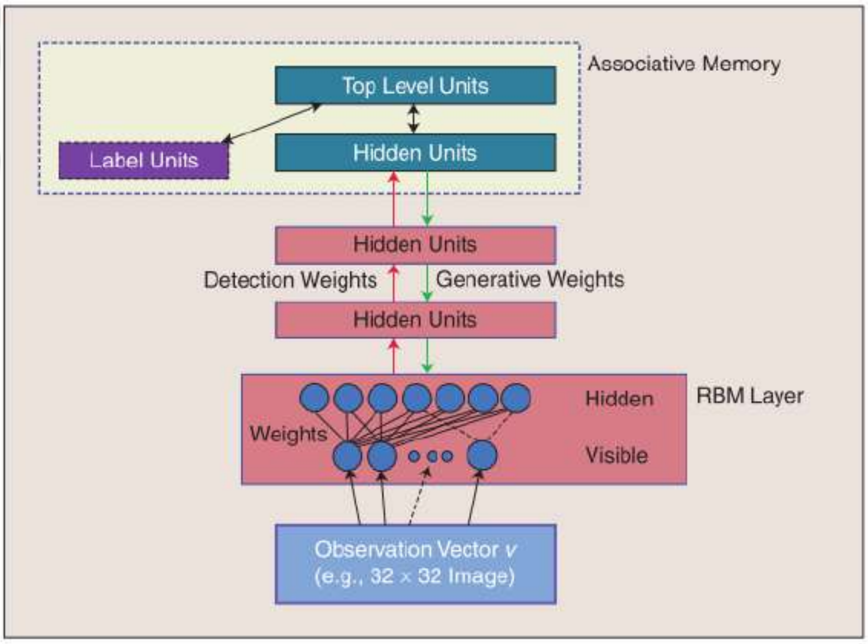


**Figure 5. Structure of the Deep Belief Network [4]**

The pre-training phase considers the top two layers of associative memory as RBM [1]; it occurs in an unsupervised greedy layer-by-layer manner and this phenomenon is called as contrastive divergence which is given by Hinton [13]. During this phase input, represented by v, is provided at the visible layer which is forwarded to the hidden layer. The original input is then reconstructed by statistically analyzing the visible layer input which is then forwarded to the hidden layer. The process of repeating these steps in both directions is known as Gibbs sampling [4]. The performance of DBNs is better than networks trained exclusively with back-propagation because in this case back-propagation is only required for performing local search on weight space which will increase the speed of training and convergence time [14]. DBN can also be used for performing unsupervised tasks in conjunction with continuous valued inputs as presented in [15]. Recently DBN are used with stacked auto-encoders in place of RBM [15] [16]. Recent work presented in [17] formulates Convolutional Deep Belief Network (CDBN) by combining DBN with CNN where "pooling" layer is present between every two adjacent RBM layers and this is particularly designed for images having large dimensions.

## IV. Applications

Deep learning have been successfully applied to variety of domains which shows its effectiveness. In this section, we present few application areas of deep learning and a brief summary of their work. MNIST [18] is a benchmark in pattern recognition methods and learning techniques by reducing efforts on preprocessing and formatting. Microsoft researchers optimized the performance of MNIST as shown in [19] by using two policies. First, expansion of size of training set by elastic distortions. Second, they used CNN. Face detection is also one of the application area of deep learning and its performance is increased by combining CNN and Gabor filter [20]. After the first step, four sub images are obtained after extracting facial features using Gabor filter. In the next step, CNN is applied to these four images. The other academic research areas of deep learning are natural language processing [21], and speech recognition and detection [8].

## V. Challenges

The challenges in the field of deep learning as given in [22] are:

- **Multimodality:** It includes design of a system that can map both text and images into the same representation space.
- **Black-box learning:** This challenge includes design of dataset which is not readable by human and data is not known.
- **Facial recognition:** It includes design of an algorithm that can recognize features better and faster than humans can.

## VI. Conclusion and Future Scope

In this paper, we studied two deep architectures that are CNN and DBN. We also discussed learning algorithms used in these two architectures and various application areas of deep learning. From this survey, we can conclude that learning deeper architectures is a necessity in the field of machine learning. Deeper architectures can perform better than shallow architectures in complicated AI tasks but learning parameters are high [1]. There are still many domain-specific tasks which are not improved by deep learning and a lot of work is to be done in these areas. The various examples are Natural Language Processing (NLP), in an expert system [1], reading the characters at the bottom of bank cheque [23]. Despite this, it is undoubtedly clear that deep learning will give a new direction to machine learning and artificial intelligence.

## References

[1]. Dandan Mo, “A survey on deep learning: one small step toward AI,” Dec. 2012.
Available: www.cs.unm.edu/~pdevineni/papers/

[2]. J. Schmidhuber, “Deep learning in neural networks: an overview,” Neural Networks, vol. 61, pp. 85-117, Oct. 2014.

[3]. L. Deng, Microsoft Research, “A tutorial survey of architectures, algorithms, and applications for deep learning”, APSIPA Trans. on Signal and Information Processing, Cambridge University Press, 2012.

[4]. I. Arel, D. C. Rose, and T. P. Karnowski, “Deep machine learning—a new frontier in artificial intelligence research,” IEEE Comput. Intell. Mag., vol. 5, no. 4, pp. 13-18, Nov. 2010.

[5]. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. Lang, “Phoneme recognition using time-delay neural networks,” IEEE Trans. Acoust., Speech, Signal Processing, vol. 37, pp. 328–339, 1989.

[6]. http://deeplearning.net/tutorial/lenet.html

[7]. D. George, “How the brain might work: A hierarchical and temporal model for learning and recognition,” Ph.D. dissertation, Stanford Univ., Stanford, CA, 2008.

[8]. S. Sukittanon, A. C. Surendran, J. C. Platt, and C. J. C. Burges, “Convolutional networks for speech detection,” Interspeech, pp. 1077–1080, 2004.

[9]. P. Y. Simard, D. Steinkraus, and J. C. Platt, “Best practices for convolutional neural networks applied to visual document analysis,” in Proc. 7th Int. Conf. Document Analysis and Recognition, 2003, pp. 958–963.

[10]. F. H. C. Tivive and A. Bouzerdoum, “A new class of convolutional neural networks (SICoNNets) and their application of face detection,” in Proc. Int. Joint Conf. Neural Networks, 2003, vol. 3, pp. 2157–2162.

[11]. Y.-N. Chen, C.-C. Han, C.-T. Wang, B.-S. Jeng, and K.-C. Fan, “The application of a convolution neural network on face and license plate detection,” in Proc. 18th Int. Conf. Pattern Recognition (ICPR’06), 2006, pp. 552–555.

[12]. H. Mobahi, R. Collobert, and J. Weston, “Deep learning from temporal coherence in video,” in Proc. 26th Annu. Int. Conf. Machine Learning, 2009, pp. 737–744.

[13]. G. E. Hinton, “Training products of experts by minimizing contrastive divergence,” Neural Comput., vol. 14, pp. 1771–1800, 2002.

[14]. G. E. Hinton and R. R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, vol. 313, no. 5786, pp. 504–507, 2006.

[15]. Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle, “Greedy layer-wise training of deep networks,” in Advances in Neural Information Processing Systems 19 (NIPS’06). 2007, pp. 153–160.

[16]. P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders,” in Proc. 25th Int. Conf. Machine Learning (ICML’08), 2008, pp. 1096–1103.

[17]. H. Lee, R. Grosse, R. Ranganath, and A. Ng, "Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations," in Proc. 26th Int. Conf. Machine Learning, 2009, pp. 609–616.

[18]. The MNIST database of handwritten digits [Online]. Available: http://yann.lecun.com/exdb/mnist/

[19]. P. Y. Simard, D. Steinkraus, J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," One Microsoft Way, Redmond WA 98052, 2000.

[20]. B. Kwolek, "Face detection using convolutional neural networks and Gabor filters," in Lecture Notes in Computer Science, vol. 3696, 2005, pp 551-556.

[21]. H. Lee, Y. Largman, P. Pham, and A. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in Advances in Neural Information Processing Systems 22 (NIPS'09), 2009.

[22]. http://deeplearning.net/icml2013-workshop-competition/challenges/

[23]. S. V. Rice, F. R. Jenkins, and T. A. Nartker, "The fifth annual test of OCR accuracy," Information Sciences Res. Inst., Las Vegas, NV, TR-96-01, 1996.