VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

OBJECT RECOGNITION USING THE PRINCIPLES OF DEEP LEARNING ARCHITECTURE

D. Malathi, J. D. Dorathi Jayaseeli, K. Senthil Kumar and S. Gopika
Department of Computer Science and Engineering, SRM University, Chennai, India
E-Mail: malathi.d@ktr.srmuniv.ac.in

ABSTRACT

This paper aims at applying the techniques of deep learning and study the behavior of its effective score comparing with traditional approaches like supervised learning and proposed to come up with a revised algorithm with application towards hand written character recognition using the principles of deep learning architecture and analyze its performance with the conjunction of benchmarking machine learning dataset like MNIST. Hand written character recognition is achieved using the deep learning model namely Deep Belief Network which is trained using a simple Restricted Boltzmann Machine (simple RBM) and three layers of Restricted Boltzmann Machine (stacked RBM). The performance of our model shows 92% accuracy. This shows that it outperforms the traditional supervised learning methods. This method can be extended to efficient text extraction in complex images.

Keywords: deep learning, sparse auto encoder, stacked de-noising auto encoders, restricted boltzmann machine.

1. INTRODUCTION

a) A preamble to deep learning

What is learning with respect to machine? The machine is trained so that it acquires new knowledge, or modifying and reinforcing existing knowledge, modifying behaviors, skills, values, and/or preferences. Deep learning [1][2] consist of a set of algorithms in machine learning that tries to learn layered models of inputs, called neural networks. The layers in neural network models correspond to distinct levels of concepts, where higher-level concepts are defined from lower-level ones. Deep learning is part of machine learning methods based on learning representations. An image can be represented in many ways for e.g., a vector of pixels. Some representations make it convenient to learn tasks of interest from the image for example features extracted from the image make it easier to represent an image. Research in this area strives to define what makes better representations and how to learn them. Most successful deep learning methods involve artificial neural networks. Deep Learning Neural Networks had its history from 1980. Ronan Collobert has said that "deep learning is just a buzzword for neural nets".

2. RECENT DEVELOPMENTS IN DEEP LEARNING

Dumitru Erhanet [3] has imitated high-level class-specific neurons using unlabeled data. They worked out this by combining ideas from latterly developed algorithms to learn invariances from unlabeled data. Their implementation scales to a cluster with thousands of machines. Their task proves that it is possible to train neurons to be discriminatory for high-level concepts using wholly unlabeled data. In the experiments, they received neurons that function as detectors for human faces, bodies, and cat faces by training on random frames of YouTube videos. These neurons instinctively capture complex invariances such as out-of-plane and scale invariances. The learned neural network representations [4] also work well for selective tasks. On the outset they obtain 15.8% accuracy from these representations for object recognition on ImageNet with 20,000 categories, a notable leap of 70% improvement over the state-of-the-art.

The Microsoft Researcher in the paper [5] has detailed about various possible classes of Deep architectures and where they can be applied and future directions are also elaborated. The successfulness of any machine learning algorithm which depends on representation of data is analyzed by the authors in paper [6]. In this tutorial [7], an introduction to deep learning research is given at first. Second, an analysis and summary of the huge work done in deep learning literature has been provided. How to implement the optimization principle for deep learning architectures are discussed in the article [8]. Many deep learning algorithms have produced good results in areas on vision and language data sets. From the experimental results it is observed that the best results obtained on supervised learning tasks involve an unsupervised learning component. These new algorithms made the deep models to train well but many questions are unanswered [9] like working of an unsupervised pre-trained model. In this article [10] the author introduced hierarchical neural networks for object recognition. He proposed a novel method for incrementally learning a hierarchy of features from unlabeled inputs. These deep learning algorithms together with the advancement of parallel computers, successfully tackled the problems that were impossible before, in terms of depth and input size. Kang-Won Lee [11] has proved that the problems which arise due to weighting issues from local image based approaches are eliminated by entertaining ensemble method. Based on this, their research can be applied to block image based facial expression recognition method also.

VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

3. DEEP LEARNING MODELS FOR OBJECT RECOGNITION

From these available deep learning models such as Deep Feed Forward Neural Network, Deep Belief Network, Deep Boltzmann Machine, Deep Convex Network, Kernel Deep Convex Network, Deep Stacking Network, Deep Auto-encoder Network, Deep Convolutional Neural Network, Deep Recurrent Neural Network and Long Short-Term Memory models, we have used Deep Belief Network (DBN) which was trained using simple RBM and stacked RBM.

a) Restricted Boltzmann machine

Boltzmann machine is a fully connected stochastic network that uses binary visible and hidden neurons, and stochastic update. It is suitable for storage and retrieval of binary patterns. Capacity of a Boltzmann machine to store patterns is higher than the capacity of a discrete Hopfield model. Restricted Boltzmann machine is a stochastic network with no connections among visible units and among hidden units. It is used in deep belief networks and deep Boltzmann machines. Restricted version of Boltzmann Machine (RBM) has visible nodes and hidden nodes. The input to the model is clamped on the visible nodes. Hidden nodes are used to model the dependencies and are viewed as nonlinear feature detectors.

b) Deep belief network

Stacked RBMs are used to build a DBN as a Multilayer Feed Forward Neural Network (MLFFNN). Pre-training phase of DBN: Unsupervised training of each of the RBMs independently. Use weights from pre-training phase to initialize the weights in the MLFFNN. Fine tuning phase of DBN: Supervised training of MLFFNN using the error back propagation method

TRAINING A DBN:
Unsupervised pre-training
1. The first level RBM is trained on the data.
2. The inferred states of the hidden units are used as data for training the next level RBM.
3. Step 2 is repeated until the entire network is trained.

Supervised Fine-tuning
1. After the network is pre-trained and weights are learned, the whole network is considered as a multi-layered feed-forward neural network.
2. Supervised fine-tuning is carried out using error back-propagation method.

c) Sparse auto-encoder

A sparse auto-encoder is a model trained by unsupervised learning. It refers to a neural network with a single hidden layer, where the target values are set equal to the inputs during training [12] In other words, the auto-encoder tries to approximate the identity function. By imposing additional constraints the network can be made to learn interesting features characterizing the input. One such constraint is the number of hidden units in the network. If this is below the number of input variables the network is actually learning a compressed representation of the input. Another constraint is used here is to have the average activation of the neurons being low. This is done by using a regularization term summing the KL-divergence over all hidden neurons indexed by j.

$$KL(\rho \| \rho_j)=\rho log(\rho\rho_j)+(1-\rho)log(1-\rho_1-\rho_j) \quad (1)$$

The regularization parameter ρ defines a target mean activation and the mean activation ρ_j of hidden neuron j over the whole training datasets interpreted as activation probability. Training of deep networks is performed by Yoshua Bengio [13].

4. RESULTS AND DISCUSSION

This section discusses the results that follow the implementation of the simple RBM and the stacked RBMs. Some snapshots of the prototype system are given in Figures-1, 2 and 3.With the above depicted snapshots of different modules, Figure-1 shows a single RBM with 100 hidden nodes. Each of the hidden nodes is rendered alongside the test digit in blue. Figure-2 trains a deep belief network made up of three RBMs. It learns to match pictures with their corresponding label. It takes about 15 minutes to train. Once trained it shows around 92% of accuracy. The trained DBN is saved in a file. Figure-3 shows the snapshot of result obtained by testing MNIST Test Data through trained Deep Belief Networks.

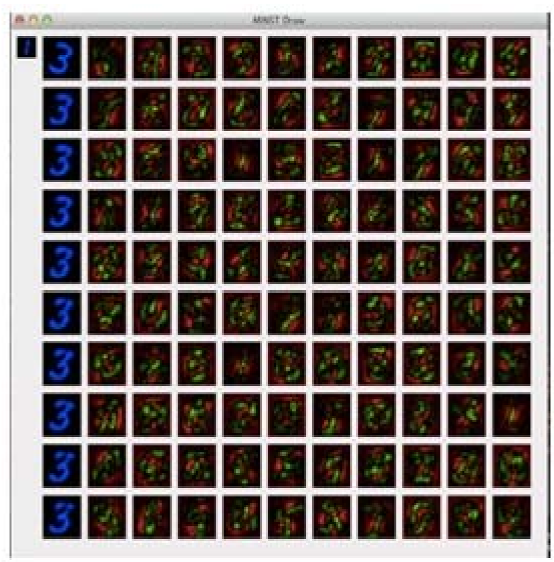

Figure-1. Simple RBM training snapshot.

VOL. 12, NO. 12, JUNE 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

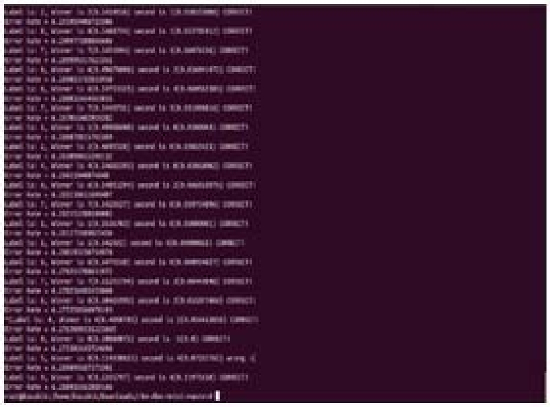

Figure-2. Stacked RBM training snapshot.

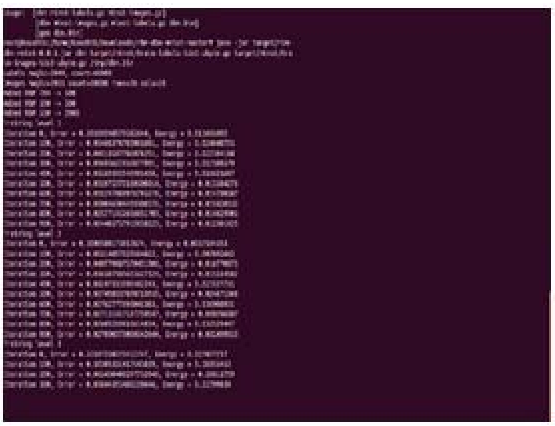

Figure-3. Testing trained DBN with MNIST Test Data snapshot.

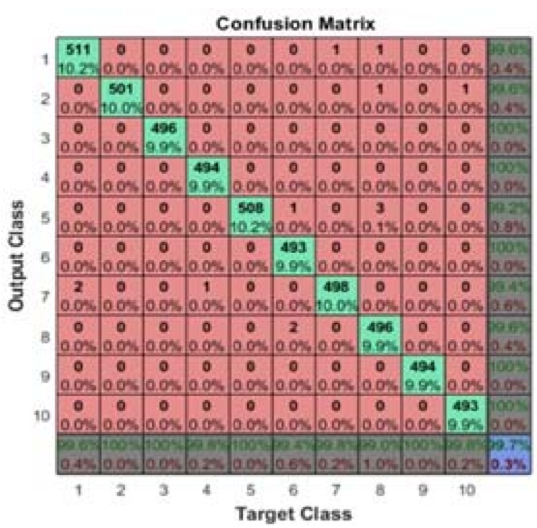

Figure-4. Confusion matrix of stacked auto encoder for handwritten digit classification using MATLAB.

5. CONCLUSION AND FUTURE WORK

From the above mentioned theoretical details of deep learning and its corresponding implementation of the theoretical details, we conclude that deep learning concepts shows a tremendous impact on the field of optical character recognitions, with the advancements of minimal learning curve as well as less time during training phase when compared to the traditional approach. Undoubtedly it has reduced the misclassification rate of the model and further this approach can be applied in the field of image-processing for its applications and natural language processing for parsing, paraphrase detection of short phrases and longer sentences and sentiment analysis.

REFERENCES

[1] Ludovic Arnold, Sebastien Rebecchi, Sylvain Chevallier, Helene Paugam-Moisy "An Introduction to Deep Learning" In Proceedings of ESANNN, 2011.

[2] Francis Quintal Lauzon, "An introduction to deep learning" in the 11th International Conference on Information Sciences, Signal Processing and their Applications: Special Sessions, 2011.

[3] Dumitru Erhan, Yoshua Bengio, Aaron Courville, Pierre-Antoine Manzagol, Pascal Vincent "Why Does Unsupervised Pre-training help Deep learning", Journal of Machine Learning, 2011 at URL: http://jlr.org/paprs/volume11/erhan10a/ erhan10a.pdf

[4] Vinod Nair, Geoffrey E. Hinton "3D Object Recognition using Deep Belief Nets" in NIPS, 2009.

[5] Li Deng, "Three Classes of Deep Learning Architecture and Their Applications: A Tutorial Survey", at URL:http://research.microsoft.com/ pubs/192937/Transactions-APSIPA.pdf

[6] Aaron Courville, Pascal Vincent, Yoshua Bengio, "Representation Learning: A Review and New Perspective" in IEEE Transactions of Pattern Analysis and Machine Intelligence, Vol.No.35, Issue8, pp.88-112, Aug'2013.

[7] Honglak Lee "Tutorial on Deep Learning and Application" in NIPS 2010 Workshop at URL: http://deeplearningworkshopnips2010.files. wordpress.com/2010/09/ nips10workshop-tutorial-final.pdf

[8] Y. Bengio "Learning deep architectures for AI" in Foundations and Trends in Machine Learning, Vol.No.2, pp.1-127, 2009.

[9] Hannes Schulz, Sven Behnke "Deep Learning: Layer Wise Learning of Feature Hierarchies", URL:http://www.ais.uni-bonn.de/ papers/ki2012 _schulz_ deeplearning.pdf

[10] Andrew Y. Ng, Jeff Dean, Greg S. Corrado, Kai Chen, Rajat Monga, Quoc V. Le “Building High-level Features using Large Scale Unsupervised Learningat URL:http://arxiv.org/pdf/1112.6209.pdf&embedded=true

[11] Kang-Won Lee , Tai-Hyung Kim , Sung-Hoon Kim , Seung-Hyung Lee , Hyon-Soo Lee, “Facial Expression Recognition using Dual Stage MLP with Subset Pre-Training”, Indian Journal of Science and Technology, Vol. 8(25), DOI: 10.17485/ijst/2015/v8i25/80978, October 2015.

[12] Pavel Vyacheslavovich Skribtsov and Pavel Aleksandrovich Kazantsev “Multi-Layer Neural Network Auto Encoders Learning Method, using Regularization for Invariant Image Recognition”, Indian Journal of Science and Technology, Vol 9(27), DOI: 10.17485/ijst/2016/v9i27/ 97704, June 2016.

[13] Yoshua Bengio, Pascal Lamblin, Dan Popovici, Hugo Larochelle “Greedy Layer –Wise Training of Deep Networks” at URL http: //machinelearning.wustl.edu/mlpaers/paper_files/ NIPS2006_739.pdf