Augmented model of stacked autoencoder for image classification

S N Shivappriya, DivyaRaju, R. Harikumar

Abstract: Stacked Auto Encoder (SAE)is used to pre-train the deep network in the training phase of the individual layer for classifying complex real time data's. MNIST and IMAGENET are used to train the network. Time consumed and accuracy during the training period is calculated for the MNIST data set which is binary image and IMAGENET dataset includes color image applying the Stacked Auto Encoder algorithm which is trained one layer at a time. Here the SAE consists of three layers which is stacked together and its parameters are varied in such a way that the constructed SAE out performs achieving time and accuracy tradeoff. The SAE model improves the accuracy of the image classifier in both binary and color image dataset with the reduced time.

Keywords-Artificial Neural Network, Stacked, Auto Encoder, Image classification

I. INTRODUCTION

An solution to the problem of "back propagation without a teacher" was first attempted by Hinton in the1980s and Rumelhartin1986[1]. Auto encoders were proposed as a solution to this problem, and the input data was used as the teacher. The back propagation algorithm is traditionally used to train neural networks and literally "back propagates" the error in the network backward to the input layers neuron from the output layers neuron, which comes under the supervised neural Network category. The capability of Deep Neural Networks is limited due to weaknesses of this algorithm. Back propagation was not effective in deep layers and also most of the available input data was unlabeled [2]. In 2006,Hint developed Deep Belief Networks(DBN) to overcome the above problems, this network composed of a stack of Restricted Boltzmann Machines (RBM) [3]. Greedy layer-by-layer training is one of the core concepts of DBNs. A similar strategy is used by stacking auto encoders yielding similar results.

High dimensional data is difficult to store and makes classification and visualization difficult. Reducing the dimensionality of such details the key to such problems. Auto encoders are alternatively known as "Auto associator" or "Diabolo networks" [4,5].Huge dimensional data or feature leads to higher complexity in the network operations, to over come this the incoming datas to the input layers are encoded and these will be reconstructed in the output layer. These auto encoders are stacked on each other to frame a deep neural network. Hence the name Stacked Auto encoder.

This multiple layers of auto encoders is widely used in reducing dimensionality[6]. Layer-wise training and tuning takes place by back propagation.

Stacked auto encoder consist of several layers of auto encoders encoding and decoding the input leading to better representational learning. Better representation refers to "the one which yields a good classifier"[7].Retaining most of the information from the input is a measure of a good representation since it is essentially the reconstruction of the input. Finally, the SAE consists of a classification layer, which classifies the images. In the current paper, we have implemented the basic model and introduced several layers with multiple nodes. The better performing models are listed in the experimental section. The structure of the auto encoder is similar to the multilayer perceptron (MLP) with the condition of the number of neurons in the input layer should be equal to the output layer neurons. Therefore, auto encoders are unsupervised learning models. Using stacking and Restricted Boltzmann Machine(RBM)[8,9] the speed of the training process gets improved. Back propagation canal so be used to train the auto encoders[10].

Generally, auto encoder[11] comes under the category of ANN with unsupervised learning. Denoisingautoencoder, sparse auto encoder, variation auto encoder, contractive auto encoder are used

In order to improve ability of getting differencing information from the data, different types of auto encoders are used such as (i)Denoising auto encoders[12] the noise components present in the input are removed in encoding process.(ii)Sparse auto encoder [13]is used when there is a need for spars it on the hidden units.(iii)Variational auto encoder[14] encodes the hidden information or buried information from the data.(iv)Contractive auto encoder[15] used to find the slight changing information from the data. In this paper among the four types of auto encoder the sparse auto encoder is selected as it is best in solving classification problems and performance of the classifier is evaluated using MNIST dataset and IMAGENET data set and its output is compared with time and accuracy trade off.

II. METHODOLGY

In this paper first the hidden layers of auto encoders are trained individually in an unsupervised manner and then train a final soft max layer, and append the layers together to frame a deep network, finally it is trained in a supervised fashion.[19-21]The soft max classifier is used in this paper as it consists of a very simple model and is there for every fast to train and predict.

Revised Manuscript Received on December 08, 2018.
SN Shivappriya,Electronics and communication engineering, Kumaraguru College of Technology,Coimbatore, Tamil Nadu, India
DivyaRaju, Electronics and communication engineering, Kumaraguru College of Technology, Coimbatore,Tamil Nadu, India
R.Harikumar, Electronics and communication engineering, Bannari Amman Institute of Technology, Sathyamangalam, Tamil Nadu ,India

A. Supervised Learning And Unsupervised Learning

In the unsupervised learning without the class label the datas are grouped, with the buried information behind the data.The main goal of unsupervised learning is to find the data structure or clusters in the data.

In the supervised learning with the class label training of the data has been takes place, but in the unsupervised learning there is no need of class information for the data.

The supervised learning functions as follows:

1. The network is trained with handwritten digits and colour images, which is collected from real time.
2. While training process input data's are converted into feature vectors with vivid features of the original data.
3. Based on the features the datas are placed on to the correct class, accordingly the performance of the classifier is computed.

B. Training Process of Auto Encoder

The presence of hidden layers represents the tuning task for training a large network. This is attributed to two problems:

(i) Poor local minima: As the depth increases with the addition of layers,there is an increased probability of finding poor local minima. The present paper implements regularization as a solution to this problem. (ii)Vanishing gradient problem:Vanishing Gradient Problem is known as "diminishing gradient flow" or "long time lag".The neurons in the higherlayers pass on the errors to the lower layers.Based on the errors in the output layers,the weights are updated directly in the higher layer. As sub sequent layers are added,they are updated based on the error in the higher layer.Therefore, decay is observed in the error rate.Greedy Layer WiseTraining was introduced as a solution to the vanishing gradient problem.Layer wise training in volves two stages:

a) Pre-Training: In this step, one layer is trained at a time using unsupervised learning. Use of labeled data makes this phase more appealing since this is supervised. The representation from each layer serves as input for the next layer, and a new representation is learned which can predict the variables of interest. By training one layer at a time, fewer local minima are involved and also cleaner gradients. This optimization strategy also helps by initializing the weights in a region of the good local minimum. With the fine-tuning step the difficult problems are over comed. As the pre-training progresses through the different layers, the previous layers from which the representation output serves as the input are greedily and conveniently ignored.

b) Fine tuning: The greedy layer supervised strategy provides a good initialization step for fine-tuning. Then in the unsupervised pre-training step, in order to improve classification process, optimization of the lower level features has been derived with supervised fine-tuning.Stochastic Gradient Descent is used to optimize the loss function by calculating the gradients and fine-tune the network using back propagation.

The autoencoder plays a vital role in classifying the larger dataset, because of its curse in dimensionality [16].For encoding and decoding the distribution of data the auto encoders use neural networks. The training process of an auto encoder consist of the following:

(a) With the different back propagation(conjugate gradient method, steppest descent, etc.) methods the auto encoder is trained.

(b) In the Auto encoder, the first few layers become in significant once the errors are back propagated to them because the network will almost learn to reconstruct the average of all the training data. With the pre training technique the above problem is solved with the use of initial weights that approximate the final solution.

(c) Initially, single auto encoder is constructed in a similar way two more layers are appended, which frames three layer spare auto encoder.

III. DATASET

In the field of machine learning the data sets are an integral part, which should chosen properly before going to the classification problem. The progression in learning algorithm plays a major role in recent days machine learning techniques. In this work two different datasets MNIST data set and IMAGENET dataset are used for the classification problem. The data sets are publicly available in order to compare the accuracy of different approaches.

A. MNIST Data Set

For digit classification the Modified National Institute of Standards and Technology database was introduced [17]. Here the images are composed 10 classes of handwritten digits with 28×28 Pixel size, which contains10 digits that ranges from 0 to 9. Previously LeNet-5 dataset is used for hand written digits classification the accuracy is lower than the MNIST dataset.

B. IMAGENET Data Set

For visual object recognition software research [18] the IMAGENET database was created. In this paper ALEXNET is used which is a subset of the IMAGENET database belonging to IMAGENET Large- Scale Visual Recognition Challenge(ILSVRC).The accuracy and time for apart of IMAGENET is calculated using the SAE algorithm.

IV. RESULT AND DISCUSSION

Twenty random images from the MNIST data set as shown in Fig1 is taken and it is classified using stacked auto encoder where its time and accuracy is calculated. Another 20 images from IMAGENET data set is taken and it is classified using the same stacked auto encoder for which its time and accuracy is calculated and the results obtained from these two data set using the same constructed auto encoder are compared where higher accuracy is obtained.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

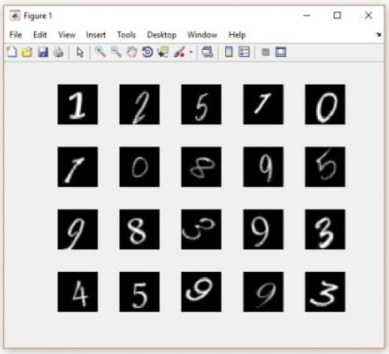

Fig1.Classified MNIST images using Stacked Auto-Encode

These set of images are iterated 50 times where the time and accuracy are measured where the time is increased and the loss is tried to be reduced during each iteration as shown in Fig2-11.Here the iteration refer so the repetition of a process. Lesser the execution time, higher the processing speed, which is the power of that network. Accuracy and loss is another important parameter accordingly the learning model should be framed with minimum loss function. If the learning model is framed by considering all the parameters 100% accuracy for predicting correct class can be achieved.

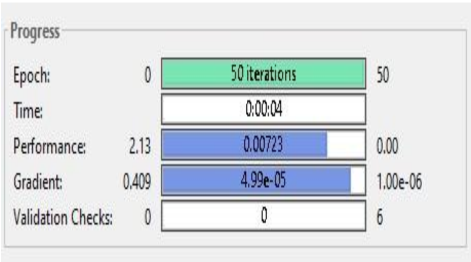

Fig2.50iterations

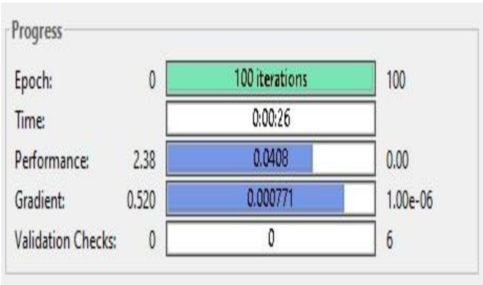

Fig3.100iterations

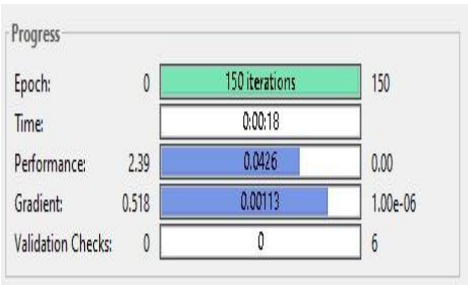

Fig4.150iterations

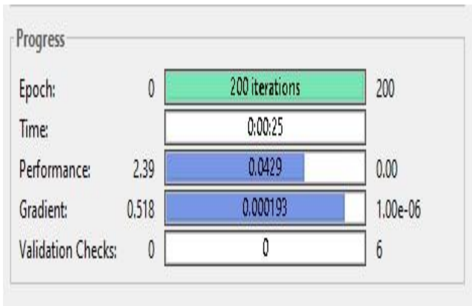

Fig5.200iterations

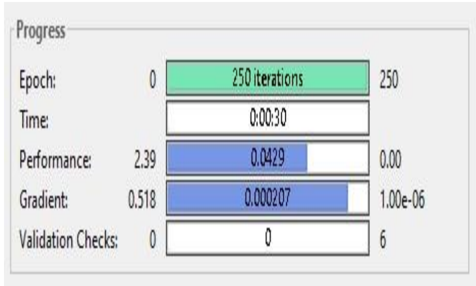

Fig6.250iterations

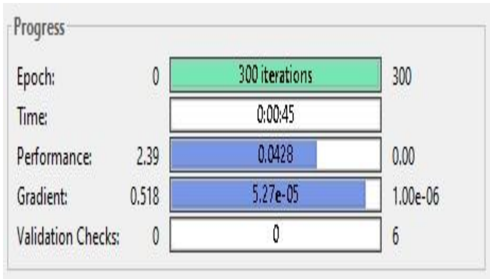

Fig7.300iterations

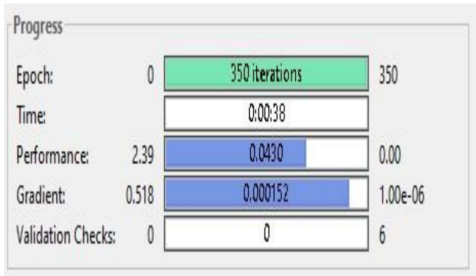

Fig8.350iterations

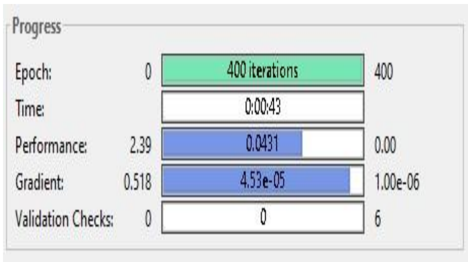

Fig9. 400iterations

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

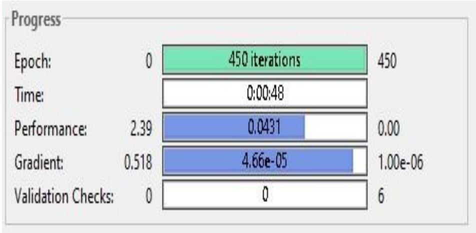

Fig10.450iterations

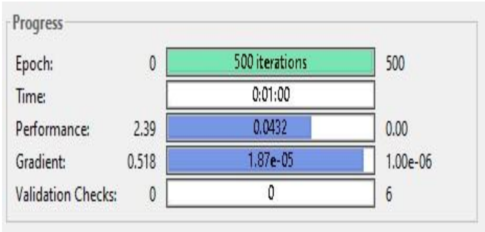

Fig11.500iterations

Fig12.Classified color image using SAE

From the IMAGENET dataset, the set of 20 images are iterated for 500 times. During 50th iteration itself 100% accuracy is obtained but when the numbers of iterations are increased the loss is increased slightly because the search for the target value is deviated. A set of color images are taken as shown in Fig12 which belongs to a set off our class i.e., Math work cap, Mathwork cube, Mathwork playing card,Mathworks crew driver as shown in Fig13 and it is trained using the same constructed SAE and the time and accuracy is calculated. Here the time to train the colour images is increased when compared to the binary image as there are more dimension in colour images when compared to binary image.

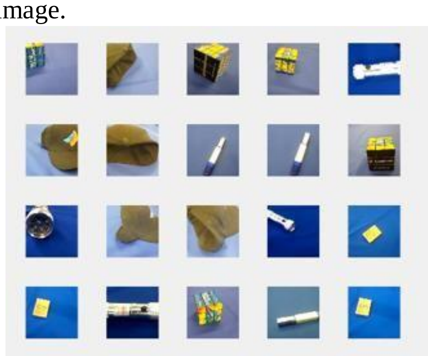

Fig13.Colorimageclassws

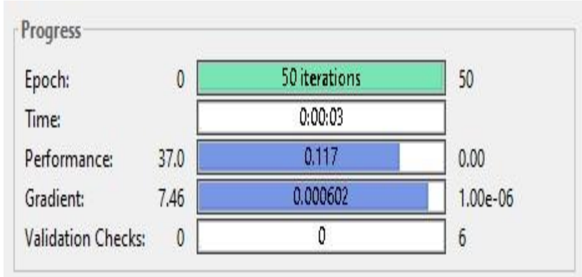

Fig14.50iterations

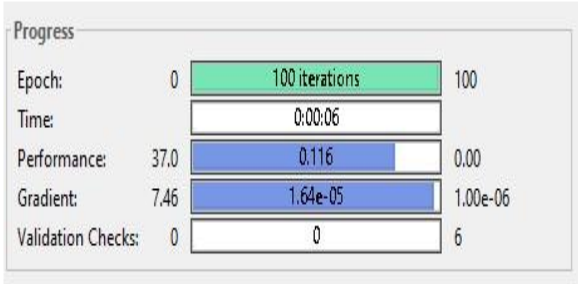

Fig15.100iterations

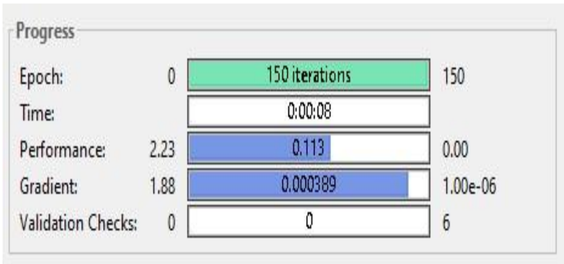

Fig16.150iterations

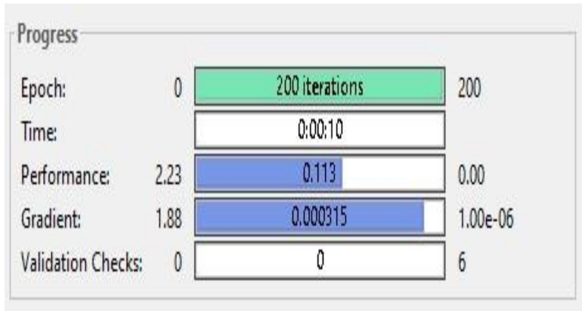

Fig17.200iterations

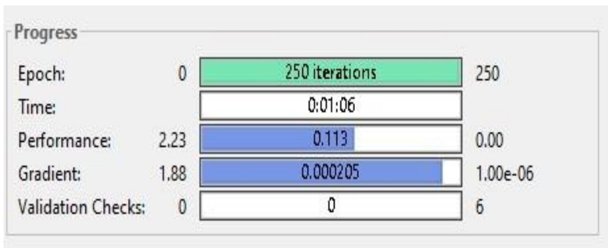

Fig18.250iterations

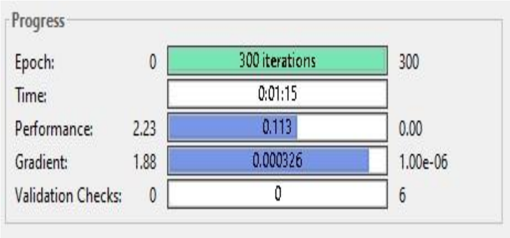

Fig19.300iterations

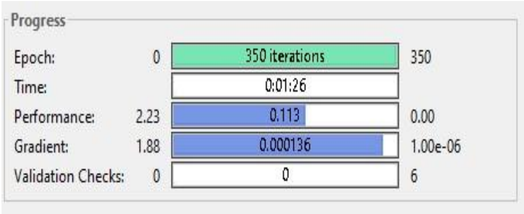

Fig20.350iterations

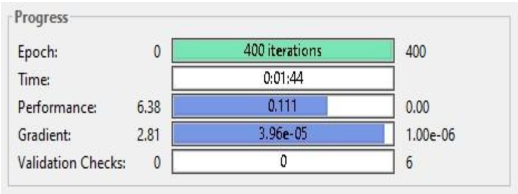

Fig21.400iterations

For binary image the time required to achieve 100% accuracy is only 4seconds and the loss is also minimum whereas for color image the loss is high as the color image has higher dimension and requires more time and hence the accuracy achieved is only 98% but the time required is only 8 seconds.

Nowadays Deep Neural Networks are widely used, instead the SAE can also be used as it achieves similar trade-off between speed and accuracy.

V. CONCLUSION

Never the less, the evolution of new concepts over the decades has led to improvements over the earlier proposed learning methods. The increasing demand for newer technologies and their cross- functionality across different domains has also catapulted to an altogether new level. Even though the technologies are improving in a faster manner, but the infrastructure is not developing in that fast. The growing claim for Stacked Auto Encoder (SAE) has brought it altogether to a new level. This paper shows that the SAE can outperform in terms of achieving the speed and accuracy trade off when test educing MNIST dataset and IMAGENET dataset. It's how that to process a set of binary and colour images the time required is very less i.e., it requires only 4 and 8 seconds. This proves the effectiveness of fusing SAE layer by layer.

REFERENCES

1. P.Baldi,"Autoencoders, Unsupervised Learning, and Deep Architectures," ICML Unsupervised Transf. Learn., pp. 37–50,2012.
2. D.Meyer,"IntroductiontoAutoencoders,"pp. 1–8,2014.
3. Y.Bengio,P.Lamblin,D.Popovici,andH.Larochelle,"Greedy Layer-Wise Training of Deep Networks," Adv.Neural Inf. Process.Syst.,vol.19, no. 1,p. 153,2007.
4. Y. Bengioetal., Learning Deep Architectures for AI, vol.9, no. 1.2011.
5. H.Larochelle, D.Erhan, A.Courville, J.Bergstra, and Y. Bengio, "An empirical evaluation of deep architectures on problems with many factors of variation," Proc.24th Int.Conf.Mach.Learn.-ICML'07,no.January,pp.473–480, 2007.
6. G.E. Hinton and R.R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," Science(80).,vol.313, no. 5786, pp. 504–507,2006.
7. P. Vincent PASCALVINCENT and H.Larochelle LAROCHEH, "Stacked Denoising Auto encoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion Pierre-Antoine Manzagol,"J. Mach.Learn.Res.,vol.11, pp. 3371–3408,2010.
8. Y.Nomura, A.S.Darmawan, Y.Yamaji, and M.Imada, "Restricted-Boltzmann- Machine Learning for Solving Strongly Correlated Quantum Systems," no. 1, pp.1–9, 2017.
9. G.C. Pacheco, R. A. Krohling, and C.A.S.daSilva, "Restricted Boltzmann machine to determine the input weights for extreme learning machines," Expert Syst.Appl.,vol.96, pp. 77–85, 2018.
10. S.Billings, "Gradient Augmented Information Retrieval with Auto encoders and Semantic Hashing,"2018.
11. A.Galloway, G.W.Taylor, and M.Moussa, "Predicting Adversarial Examples with High Confidence," no. Icml,2018.
12. G.Du,L.Yuan,K.J.Shin, and S.Managi, "Enhancement of land-use change modeling using convolutional neural networks and convolutional denoising auto encoders," pp.1–26, 2018.
13. A.Sankaran, M.Vatsa, R.Singh, and A.Majumdar, "Group sparse auto encoder," ImageVis.Comput., vol. 60, pp. 64–74,2017.
14. W. Jin, R. Barzilay, and T.Jaakkola, "Junction Tree Variational Auto encoder for Molecular Graph Generation,"2018.
15. D.Kohli,B. C.Das, V.Gopalakrishnan, and K.N.Iyer, "Learning rotation in variance in deep hierarchies using circular symmetric filters," ICASSP, IEEE Int.Conf.Acoust. Speech Signal Process. - Proc., pp.2846–2850, 2017.
16. R.Iyer and M.Lewis, "Transparency and Explanation in Deep Rein for cement Learning Neural Networks," 2018.
17. G.Cohen, S.Afshar, J.Tapson, and A.VanSchaik, "EMNIST: Extending MNIST to handwritten letters," Proc.Int.Jt.Conf. Neural Networks, vol. 2017–May, pp.2921–2926, 2017.
18. P.Chrabaszcz, I.Loshchilov, and F.Hutter, "A Down sampled Varian to fImageNet as an Alternative to the CIFAR data sets," pp.1–9, 2017.
19. S. Saravanakumar, V. Dinesh Kumar, "High Throughput Quaternary Signed Digital Adder Design For Portable Electronic Applications", International Journal of Pure and Applied Mathematics, Volume 116 No. 11 2017, 61-69.
20. T. Karuppusamy, C. Velmurugan, S. Saran, K. Sukanthanbabu, "Investigation On The Microstructure And Wear Characteristics Of Heat Treated Hybrid Aluminium Composites", Journal of Advanced Research in Dynamical and Control Systems, Vol. 9. Sp– 6 / 2017, pp1895-1912.
21. S.Gayathiri, R.Gokulapriya,V.S.Arulmurugan," An Improved Channel Aware Smart Grid Transmission For Manet", International Journal of Innovations in Scientific and Engineering Research (IJISER), Vol 1 Issue 4 APR 2014, pp212-219.

*Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication*