A Study on Deep Neural Networks Framework

Huang Yi
Department of Electronic and Optical Engineering
Ordnance Engineering College
Shijiazhuang, China
dd_huangyi@163.com

Duan Xiusheng
Department of Electronic and Optical Engineering
Ordnance Engineering College
Shijiazhuang, China
sjzdxsh@163.com

Sun Shiyu
Department of Electronic and Optical Engineering
Ordnance Engineering College
Shijiazhuang, China

Chen Zhigang
68129 Army
Lanzhou, China
285759305@qq.com

***Abstract*—Deep neural networks(DNN) is an important method for machine learning, which has been widely used in many fields. Compared with the shallow neural networks(NN), DNN has better feature expression and the ability to fit the complex mapping. In this paper, we first introduce the background of the development of the DNN, and then introduce several typical DNN model, including deep belief networks(DBN), stacked autoencoder(SAE) and deep convolution neural networks(DCNN), finally research its applications from three aspects and prospects the development direction of DNN.**

Keywords*—deep neural networks; deep belief networks; stacked autoencoder; deep convolution neural networks

I. Introduction

Deep Neural Networks(DNN) is an important method for machine learning and has been widely used in many fields. DNN is inspired by structure of mammalian visual system. Biologists have found that mammalian visual system contains many layers of neural network. It processes information from retina to visual center layer by layer, sequentially extracts edge feature, part feature, shape feature and eventually forms abstract concept [1].

In general, the depth of DNN is greater than or equal to 4. For example, a Multilayer Perceptron (MLP) with more than 1 hidden layer is a DNN framework. DNN extracts feature layer by layer and combine low-level features to form high-level features, which can find distributed expression of data. Compared with shallow neural networks(NN), DNN has better feature expression and ability to model complex mapping[2]. However, due to the gradient diffusion problem caused by depth of DNN, DNN can't carry out effective training. To solve this problem, in 2006, G. E. Hinton, the professor of the University of Toronto, proposed a weight initialization method. By unsupervised pre-training RBM layer by layer and initializing DNN with the weight learnt, the problem of training DNN got an effective solution[3]. DNN enters into a flourishing period. Through the introduction of several typical DNN models and its applications in various fields, readers can have a general understanding of DNN and its current study situation.

II. Several Typical DNN Model

Frequently-used DNN model are mainly three types: Deep Belief Networks(DBN), Stacked Aoutoencoder(SAE),Deep Convolution Neural networks(DCNN).

A. Deep Belief Network(DBN)

DBN is a DNN model proposed by G. E. Hinton in 2006. It has no difference with MLP from the network structure, but DBN uses greedy layerwise algorithm based on RBM to initial weight matrix of DBN while MLP initials weight matrix randomly. Its training procedure is as follows [3]:

1) *Step 1:* Pre-training

a) make two adjacent layers as RBM and split DBN into several RBMs, as shown in figure1 pretraining.

b) start from the bottom RBM and train it.

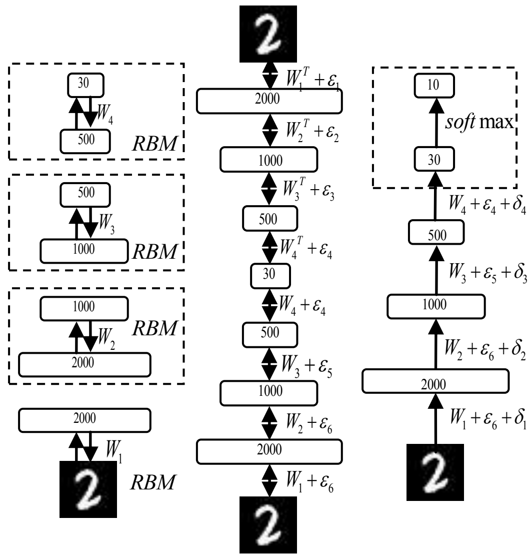
c) use the last RBM's output as the next RBM's input and train the next RBM.

d) repeat c) until all RBM training well, save weight matrix W.

2) *Step 2:* Fine-tuning

a) construct a reconstruction procedure for DBN, initialize it with W^T matrix, as shown in figure1 finetuning.

b) finetun DBN with BP algorithm using reconstruct error as object function, save new weight matrix W+e.



Pretraining *Finetuning* *Classification*

978-1-4673-9613-4/16/$31.00 ©2016 IEEE

Fig. 1. Training procedure of DBN.

After pre-training, by adding SVM, softmax to the top of DBN, it can be used for classification, prediction, regression tasks and so on. G. E. Hinton demonstrated DBN on MNIST handwriting recognition task and verified the DBN superior to feedforward neural networks[3]. Since DBN's proposal, researchers do many improvements to it, such as Convolutional Deep Belief Networks(CDBN)[4], Sparse Deep Belief Networks(SDBN) [5]and so on.

B. Stacked Aoutoencoder(SAE)

On the basis of studying DBN, in 2007,Y. Bengio et al. proposed Stacked Autoencoder(SAE). Different with DBN, SAE is based on AE instead of RBM[6]. Their training procedure is as follows[6]:

1) *Step 1:* Pre-training

a) make two adjacent layers as AE and split SAE into several AEs, as shown in figure2 pretraining.

b) start from the bottom AE and train it by minimizing its reconstruction error.

c) use the last AE's output as the next AE's input and train the next AE.

d) repeat c) until all AE training well, save weight matrix W.

2) *Step 2:* Fine-tuning

a) construct a reconstruction procedure for SAE, initialize it with W^Tmatrix, as shown in figure2 finetuning.

b) finetun SAE with BP algorithm using reconstruction error as object function, save new weight matrix W+e.

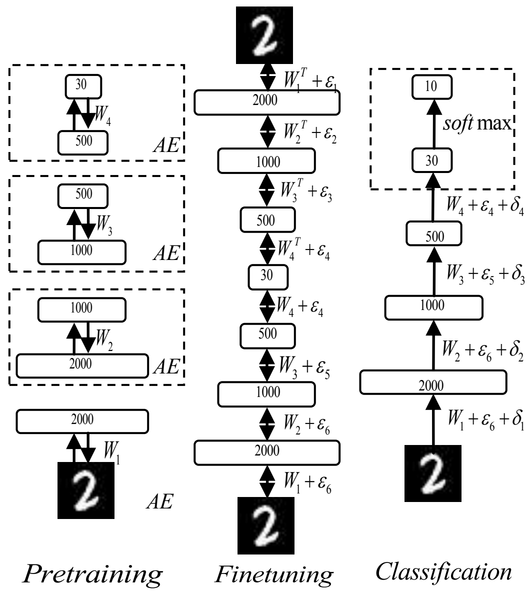


Fig. 2. Training procedure of SAE.

The same with DBN, SAE is trained with greedy layerwise algorithm, but it's easier to train AE than RBM because AE uses reconstruction error as object function[6]. The same with the RBM, AE is an unsupervised self-learning neural network, so we can construct AE layer by layer to train it. In 2008, H. Lee et added sparse constraint to the object function, proposed Sparse Autoencoder(SAE), which can learn sparse feature expression [7]. P. Vincent put forward Denoising Autoencoder(DAE), which mixed input data with some noise to improve the robustness of the network [8]. 2014, Wei Wang put forward Generalized Autoencoder(GAE), by utilizing relation between the different dimensions of the input data, achieved good result [9].

C. Deep Convolution Neural network(DCNN)

Convolution Neural networks(CNN) was proposed by Yan Lecun in 1989. It is a DNN model that can use BP Algorithm directly. A single layer CNN contains convolution, nonlinear transformation and pooling three phases.

Convolution: convolution kernel detect all positions of the input feature map and realize the weight sharing in the same input feature map. In order to extract different features from input feature map, different convolution kernels are used. For $n_1 \times n_2$ input x, using m_1 convolution kernels of $m_2 \times m_3$ to convolute it, obtain $(n_1 - m_2 + 1) \times (n_2 - m_3 + 1)$ feature map y. Its expression is as follows.

$$y_i = b_i + W_i * x \tag{1}$$

Among them, $i \in [1, m_1]$; $*$ is the two-dimensional discrete convolution operator; b is the bias parameter.

Nonlinear transformation: in order to screen features from the feature map of the convolution phase, nonlinear transformation is usually chosen. Using The feature map of the convolution phase as input, do nonlinear mapping as follows.

$$R = H(y) \tag{2}$$

Among them, y is the convolution output; H is the mapping function; R is output of Nonlinear transformation.

Commonly used mapping function are sigmoid, tanh, ReLU, etc.

Pooling: operate independently for each feature map, usually using the average pooling or maximum pooling. The main function of this operation is to reduce the resolution and the calculation cost.

The weight sharing of CNN greatly reduced the number of parameters and the training cost of the network, and using the BP algorithm directly without pretraining greatly accelerated the network's training efficiency. We can get deep convolution neural networks(DCNN) by piling up CNN.

III. REVIEW ON APPLICATION OF DNN

A. Speech recognition

In the traditional speech recognition technology, HMM-GMM (Gaussian Mixture Model) is widely used. because it's an shallow model, with the development of DNN, more and more people choose to use DNN. a Mohamed [10-12] applied DNN to acoustic model, using DNN model instead of the GMM model, in a variety of speech recognition task consistently achieved better result than HMM-GMM model.

D Yu proposed the HMM-GMM-BN framework[13-15], which used DNN to extract bottle neck(BN)feature parameters to replace traditional HMM-GMM speech feature parameters and train HMM-GMM. Experimental results show that the speech recognition system based on HMM-GMM-BN framework can obtain performance compared to HMM-DNN. In 16-17, rectified linear units(ReLUs) activation function was used to replace the sigmoid activation function in HMM-DNN. The results show that HMM-DNN with ReLUs get better performance. O Abdel applied CNN to acoustic modeling for speech recognition. It improved the robustness of different speaker in acoustic mode[18-19]. K He combined CNN with ReLUs and use it to acoustic modeling, improved the performance of recognition[31].

Z Ling proposed HMM-RBM and HMM-DBN for speech synthesis[20]. This method can preserve the spectral details and reduce the over-flatten of the synthesized speech. Aiming at the characteristics of speech synthesis, H Zen proposed Multi-distribution deep belief network(MD-DBN) [21], which is a speech synthesis method based on DBN. With different types of RBM in MD-DBN, it can model spectrum/pitch feature and clear/muddy feature and estimate the joint probability distribution of the syllable and acoustic features. Y Wang presented a speech enhancement method based on ideal binary time-frequency masking estimation[22], which transform a speech enhancement problem into a classification problem of estimating ideal binary time-frequency masking estimation using DNN.

B. *Face recognition and ILSVRC*

Face recognition is widely used in department like government, military and bank for security and has a great prospect.

Typically, the DeepID [23]of The Chinese University of Hong Kong and the DeepFace[24] of Facebook got a correct rate of 97.45% and 97.35% respectively in labeled faces in the wild(LFW) database, only slightly lower than the correct recognition rate of human 97.5%. Then DeepID2[25] improve the performance to 99.15%, better than all deep learning algorithm or non deep learning algorithm.

ImageNet large scale visual recognition challenge (ILSVRC) is a international competition to evaluate the large-scale image classification and object recognition algorithm. Every year, there are many participating teams from company, research institute and University.

In 2012, A.Krizhevsky first applied CNN to ILSVRC and got first in image classification and object localization task[26]. In ILSVRC-2013, M.D.Zeiler improved the DCNN proposed in [1], achieved first in image classification task[27].

In the ILSVRC-2014, almost all of the teams used CNN or its deformation[]. The GoogLeNet team conbine CNN with Hebbian theory, got first in specified data graph classification group with a classification error of 6.7%; in multi-target detection task, the NUS team used improved CNN network in network(NIN), got first with a mean average precision 37%[].

In ILSVRC-2015, inspired by highway network, team from Microsoft Asia Research Institute(MSRA) proposed a improved RESNET, by constructing a 152 layers RESNET got first in three task[30]. From DNN was first used in ILSVRC and achieved good results, to 2014, 2015 almost all teams used deep learning method, we can see DNN's great advantages compared traditional feature extraction method in the image recognition.

C. *Application in else area*

DNN is also widely used in other fields. Li Shuai presented a method to apply DNN to synthetic aperture radar occluded image feature extraction, improved the correct recognition rate by utilizing comprehensively local and global structure information of the target[31]. Cao Lei applied the DNN to the satellite orbit prediction, and established a hybrid model based on DNN to improve the accuracy of the satellite orbit prediction[32]. CunLi Mao applied DNN to the field of nonferrous metal entity recognition field, using denoising autoencoder (DAE) to train DNN layer by layer pre training, an solved the feature extraction problem of recognition task effectively in nonferrous metal entity field[33].

IV. SEVERAL TYPICAL DNN MODEL

In this paper, the background of the development of DNN, several typical DNN model and the application of DNN are summarized. The development of DNN has promoted the development of each field related to machine learning, but the development of DNN is still in its infancy, and many problems have not been solved yet.

Better DNN structure: traditional DNN is a pile of single layer neural network. [30]proposed Resnet, modified DNN structure to obtain a good result. The structure of DNN has great influence on performance of DNN.

Better algorithm: the greedy algorithm is essentially a kind of weight matrix initialization method, and the core of the algorithm is still dominated by BP, the limitations of the BP algorithm still exist.

References

[1] T. Serre, G. Kreiman, M. Kouh, et al., “A quantitative theory of immediate visual recognition,” Progress in Brain Research, vol 165, pp. 33-56, 2007.

[2] Y. Bengion, O. Delalleau, “On the expressive power of deep architectures,” Proc. of the 14th International Conference on Discovery Science. Berlin: Springer-Verlag, 2011, pp. 18-36.

[3] G. Hinton, S. Osindero, Y The. “A fast learning algorithm for deep belief nets,” Neural Computation, vol. 18(7), pp. 1527-1554, 2006.

[4] H. Lee, H. Grosse, R. Ranganath, et al., “Ensrupervised learning of hierarchical representations with Convolutional Deep Belief Networks,” communication of ACM, vol. 54(10), pp. 95-103, 2011.

[5] X. Halkias, S. Paris, H. Glotin, “Sparse penalty in deep belief networks: using the mixed norm constraint,” http://arxiv.org/pdf/1301.3533.pdf.

[6] Y. Bengio, P. Lamblin, D. Popovici, et al., “Greedy layer-wise training of deep networks,” proceedings of 20th annual conference on neuranl

information processing systems, wancouver:neural information processing system foundation, 2007, pp. 153-160.

[7] H. Lee, C. Ekanadham, and A. Ng, "Sparse deep belief net model for visual area V2," Advances in Neural Information Processing Systems (NIPS'07). Massachusetts: MIT Press, 2008, pp. 873–880.

[8] P. Vincent, H. Larochelle, Y. Bengio. "Extracting and composing robust features with denoising autoencoders," Proc. of the 25th International Conference on Machine Learning (ICML'08). New York: ACM Press, 2008, pp. 1096–1103.

[9] W. Wei, H. Yan, W. Yizhou. "Generalized autoencoder: a neural network framework for dimensionality reduction" Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. Massachusetts: MIT Press, 2014, pp. 490-497.

[10] A. Mohamed, G. Dahl, G. Hinton. "Acoustic modeling using deep belief networks," Audio, Speech and Language Processing, vol. 20(1), pp. 14-22, 2012.

[11] A. Mohamed, G. Dahl, G Hinton, "deep belief networks for phone recognition," NIPS workshop on deep learning fir speech recognition and related application, hyatt regency eancouver, canada[s.n.],2009, pp. 1-9.

[12] G. Hinton, L. Deng, D. Yu, et al.. "Deep neural net woks for acoustic modeling in speech recognition:the shared viers of four research groups," Signal Processing Magzine, vol. 29(6) pp. 82-97, 2012.

[13] D. Yu, M. Seltzer, "Improved bottleneck features using pretrained deep neaural net works," Inter. Speech. 2011, pp. 237-240.

[14] Y. Bao, H. Jiang, L. Dai, et al., "Inchoherent training of deep neural networks to decorrelate bottleneck features for speevh recognition," ICASSP. British columbia: IEEE, 2013, pp. 6980-6984.

[15] T. Sainath, B. Kingsbury, B. Ramabhadran. "Autoencoder bottlenech features using deep belief networks," ICASSP. Kyoto: IEEE, 2013, pp. 6980-6984.

[16] G. Dahl, T. Sainath, G. Hinton. "Improving deep neural networks for lvcsr using rectified linear units and dropout," ICASSP, British columbia: IEEE, 2013, pp. 8609-8613.

[17] M. Zeiler, M. Ranzato, R. Monga, et al., "On rectified linear units for speech processing," ICASSP, British columbia: IEEE, 2013, pp. 3517-3521.

[18] O abdel, a mohamed, h jiang, et al.. "applying convolutional neural networks concepts to hybrid NN-HMM model fro speech recognition," ICASSP, Kyoto: IEEE, 2012, pp. 4277-4280.

[19] T. Sainath, A. Mohamed, B. Kingsbury, et al., "Deep convolutional neural networks for LVCSR," ICASSP, British columbia: IEEE, 2013, pp. 8614-8618.

[20] Z. Ling, L. Deng, D. Yu, "Modeling spectral envelopes using restricted boltzmann machins and deep belief networks for statistical patametric speech sythesis," Audio, Speech, and Languange Processing, vol. 21(10), pp. 2129-2139, 2013.

[21] H. Zen, A. Senior, M. Schuster, "Statistical patametric speech synthesis using deep neural networks," ICASSP, British columbia: IEEE, 2013, pp. 7962-7966.

[22] Y. Wang, D. Wang. "Towards scaling up classification based speech separation," Speech, and Languange Processing, vol. 99, pp. 1-23, 2013.

[23] Y. Sun, X. Wang, X. Tang, "Deep learning facerepresentation from predicting 10000 classes," Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition. Piscataway, NJ: IEEE, 2014, pp. 1891-1898.

[24] Y. Taigman, M. Yang, M. Ranzato, et al., "Deepface: closing the gap to human-level performance inface verification," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Piscataway, NJ: IEEE, 2014, pp. 1701-1708.

[25] Y. Sun, X. Wang, X. Tang, "Deep learning facerepresentation by joint identification-verification," CoRR, 2014, pp. 1406-4773.

[26] A. Krizhevsky, I Sutskever, G. Hinton, "Imagenet classification with deep convolutional neuralnetworks," Advances in Neural Information Processing Systems. Red Hook，NY: Curran Associates，2012, pp. 1097-1105.

[27] M Zeiler, R Fergus, "Visualizing and understandingconvolutional neural networks," CoRR, 2013, pp. 1311-2901.

[28] O. Russakovsky, J. Deng, H. Su, et al., "ImageNet large scale visual recognition challenge" CoRR, 2014, pp. 1409- 0575.

[29] M. Lin, Q. Chen, S. Yan, "Network in network," CoRR, 2013, pp. 1312-4400.

[30] K. He, X. Zhang, S.Ren, J.Sun, "Deep residual learning for image recognition," unpublished.

[31] L. Shuai, X. Yuelei, M. Shiping, et al., "New method for SAR occleded targets recognition using DNN," Journal of Xidian University, vol. 6(42), pp. 154-160, 2015.

[32] C. Lei. "Satellite orbit prediction method based on compensation model by using Deep Neural Network," Nanjing University of Aeronautics and Astronautics.

[33] M. Cunli, Y. Zhengtao, S. Tao, et al., "A kind of nonferrous metal industry entity recognition model based on deep neural network architecture," Journal of Computer Research and Development, vol. 52(11), pp. 2451-2459, 2015.