Advances in Intelligent Systems Research, volume 132
2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM 2017)

Face Recognition Based on Deep Autoencoder Networks with Dropout

Fang Li[1], Xiang Gao[2,*] and Liping Wang[3]
[1,2,3]School of Mathematical Sciences, Ocean University of China, Lane 238, Songling Road, Laoshan District, Qingdao City, Shandong Province, 266100, People's Republic of China
[*]Corresponding author

***Abstract*—Though deep autoencoder networks show excellent ability in learning feature, its poor performance on test data go against visualization and classification of image. In particular, a standard neural net with multi-hidden layers typically fails to work when sample size is small. In order to improve the generalization ability and reduce over-fitting, we apply dropout to optimize the deep autoencoder networks. In this paper, we propose face recognition based on deep autoencoder networks with dropout. Our experiments show that deep autoencoder networks with dropout yield significantly lower test error, and bring a new conception in pattern recognition with deep learning.**

Keywords-deep- autoencoder networks; dropout; face recognition

I. INTRODUCTION

Neural networks contain multiple nonlinear hidden layers and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs. However, as a typical algorithm of traditional neural networks, BP faces many problems, such as gradient diffusion and local optimal. In 2006, Hinton proposed an effective way of initializing the weights that overcomes the limitation of traditional neural networks by layer-by-layer pre-training, and constructs a framework for the later deep learning [1].

Deep networks show excellent ability in learning feature, which facilitate visualization and classification of data. Hoverer, deep networks typically perform poorly on test data when a large feed-forward neural network is trained on a small training set. This leads over-fitting. For addressing this problem, diverse methods have been advanced.

In the machine learning, the main methods of reducing over-fitting are early terminate training, data set expansion and weight penalties of various kinds. Moody proposed the moody criterion to improve the generalization ability of real-valued neural networks [2]. Nowlan and Hinton introduced weight penalties and soft weight sharing to simplify neural networks and reduce over-fitting [3]. Mackay set regularizing constants by examining posterior probability distribution [4]. A statistical theory was proposed about preventing overtraining [5]. Imrie and Durucan obtained accuracy predictions by the cascade-correlation learning algorithm [6]. Reference [7] well suppressed over-fitting based on Bayesian statistics. Ensemble improved the performance of neural networks [8]. Reference [9] presented a fully Bayesian treatment of the Probabilistic Matrix Factorization model and achieved significantly higher prediction accuracy. A Bayesian method that uses an adaptively selected number of hidden variables to combine subgroups of features into a network and improved the statistical significance of identified features [10]. However, the above methods are limited in image recognition with deep neural networks, so it is worth exploring a more effective way to reduce over-fitting of deep networks. Preventing co-adaptation of feature detectors by dropout showed performance of neural networks on supervised learning tasks in 2012 [11].

In this paper, for solving bad generalization, we model deep autoencoder networks with dropout and propose face recognition based on model in the small sample. Our experiments show that this method yields significantly lower classification error on the small facial dataset.

This paper is structured as follows. Section 2 describes theory and model of deep autoencoder networks with dropout. Section 3 discusses the application of model in face recognition. Section 4 evaluates the model. In section 5, we present our conclusion.

II. THEORY AND MODEL

Deep networks are pre-trained by restricted Boltzmann machines, autoencoder and deep Boltzmann machine without supervision. In this paper, we model a deep network with autoencoder which minimizes the discrepancy between the original data and its reconstruction by layer-by-layer pre-taining. Then the required gradients are easily obtained by using the chain rule to back-propagate error derivatives. The whole pre-training makes parameters of the model close to a good solution.

Figure 1 shows autoencoder with four layers. Assume that the first hidden layer contained 50 units; the second hidden layer contained 10 units. The network minimizes the discrepancy between the original data and its reconstruction by two encoder layers and two decoder layers. The input X generates Y_1 through the first encoder layer, then Y_1 is used as input to the next layer and creates Y_2 via the second encoder layer. Followed by, Y_2 generates Y_1' through the first decoder layer, then Y_1' is used as input to the second decoder layer and creates X'. Where W_1, W_2, W_1' and W_2' separately correspond to the encoder weights and decoder weights of deep autoencoder networks. Obviously, the whole system is unlabelled.

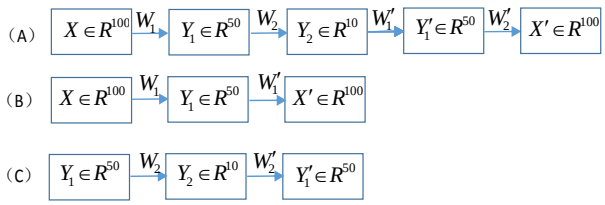

FIGURE I. PRE-TRAINING A DEEP AUTOENCODER NETWORK WITH FOUR LAYERS.

By above pre-training, we obtain approximate optimal model. What we need is Y_2 generated by the input X through two encoder layers, and establishes relationship with label information of the input X . Further, we yield optimal model of deep autoencoder networks by supervised fine-tuning, as shown in Figure 2.

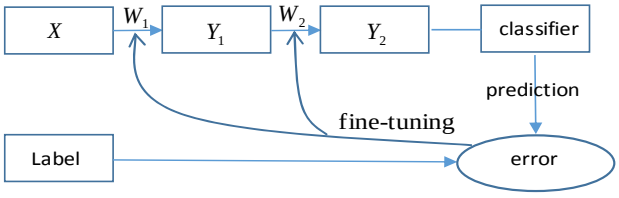

FIGURE II. FINE-TUNING DEEP AUTOENCODER NETWORKS.

The optimal model takes the visual features of facial image as input, and outputs the class information of image by connecting classifier. Though deep autoencoder networks show excellent ability in learning feature, its poor performance on test data go against visualization and classification of image. In order to improve the generalization ability and reduce over-fitting, we apply dropout to optimize the deep autoencoder networks.

A motivation for dropout comes from a theory of the role of sex in evolution [12]. The explanation for the superiority of sexual reproduction is that, sexual reproduction is not just to reduce complex co-adaptations by mix-ability of genes, but also to allow new gens to enhance their ability to fit the environment [13]. Similarly, each hidden unit in a neural network trained with dropout must learn to work with a randomly chosen sample of other units and prevent complex co-adaptations among units. This should reduce over-fitting each hidden unit on the training data [14].

Based on the above, we discover that training dropout neural nets are equivalent to adding a probabilistic process for each hidden layer as shown in Figure 3(B). Combined with probability, we need a random vector m , each of which has probability p of being 1, to sample each hidden unit. Then, m obeys independent Bernoulli distribution.

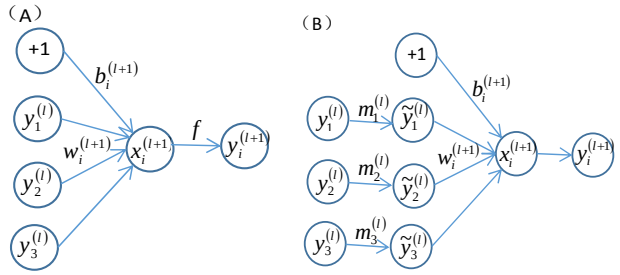

FIGURE III. COMPARISON OF THE BASIC OPERATIONS OF A STANDARD AND DROPOUT NETWORK: (A) A STANDARD NETWORK. (B) A DROPOUT NETWORK.

Consider a neural network with $L = 4$ layers. Let $l \in \{0,1,\cdots,L-1\}$ index the hidden layers of the network. Let $x^{(l)}$ denote the vector of inputs into layer l , $y^{(l)}$ denote the vector of outputs from layer l . $W^{(l)}$ and $b^{(l)}$ are the weights and biases at layer l . The feed-forward operation of a standard neural network can be described as Figure 4, where f is any activation function.

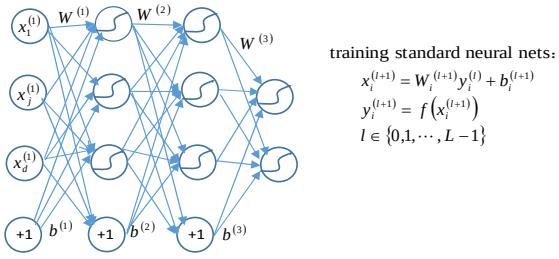

FIGURE IV. STANDARD NEURAL NET MODEL.

As previously described, random vector m samples the outputs $y_i^{(l)}$ of that layer, which create the thinned outputs $\tilde{y}_i^{(l)}$ are used as input to the next layer. By dropping a unit out, we mean temporarily removing it from the network. Each unit is retained with a fixed probability p independent of other units and the choice of which units to drop is random. This amounts to sampling a sub-network from a larger network when the process is applied at each layer. The feed-forward operation with dropout becomes Figure 5.

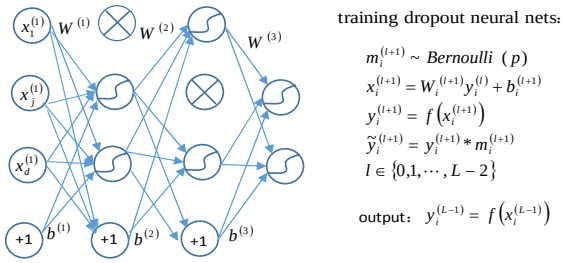

FIGURE V. DROPOUT NEURAL NET MODEL.

Similar to standard neural nets, the back-propagation of a dropout neural network can be trained using stochastic gradient

descent. The only difference is that for each training case in a mini-batch, we sample a thinned network by dropping out units. Furthermore, forward and back-propagation for that training case are done only on this thinned network.

Applying dropout to a neural net with n units can be seen as training a collection of 2^n possible thinned neural networks. These networks all share weights so that keep the number of parameters of the same size. For each presentation of each training case, a new thinned network is sampled and trained. On the test step, we ensure that for any hidden unit the expected output at training time is the same as the actual output at test time by the "mean network" that contains all of the hidden units. Just think, if a unit is retained with probability p during training, the outgoing weights of that unit are multiplied by p at test time. Otherwise, the hidden units of the actual output at test time will be more $(1-p)*100\%$ units than at training time.

III. MODEL APPLICATION

Face recognition is one of the most active and challenging research topics in computer vision and pattern recognition due to its wide-ranging applications in many areas, such as identity authentication, access control, surveillance, and human computer interaction [15]. During the past decades, considerable progress has been made in face recognition and many new methods have been proposed. For face recognition, a standard neural net with multi hidden layers fails to work when sample size is small [16]. In this paper, we pose face recognition based on deep autoencoder networks with dropout.

We achieve face recognition using the ORL database [17], which consisted of 400 facial images of 40 individuals with various facial expression and lighting directions. Each image is 92×112 pixels gray-scale image. Facial images samples in the ORL database are shown in Figure 6. The net takes the visual features of facial image as input, and outputs the class information of image by $Logistic$ function. The number of output units is the number of image classes. If the facial image belongs to $k-th$ class, the corresponding output should be that the rest are 0 except for $k-th$ unit is 1. As described above, we firstly obtain approximate optimal value by pre-training with autoencoder. Secondly, we establish contact with label information of the input image and yield optimal model of deep autoencoder networks by supervised fine-tuning $W=\arg\min_{W}\sum_{i=1}^{N}\left(f\left(I_i\right),T_i\right)$. Where I_i is the visual feature of facial image and T_i is the class label of facial image. The position of the maximum node of outputs is predicted class when test images enter deep autoencoder networks and the test error is calculated.

FIGURE VI. FACIAL IMAGES SAMPLES IN THE ORL DATABASE.

We randomly select 90% facial images as training set and the remaining 10% as test data. All images are normalized to 28 × 28. We experiment with 784-800-800-40 and use stochastic gradient descent with 40-sized mini-batches. Weights were updated at the end of each mini-batch. Figure 7 shows test error of two deep autoencoder networks. Compared with the standard deep autoencoder networks, the deep autoencoder networks with dropout (dropout rates $p=0.4$) can effectively decrease the classification error of the ORL database and enhance the generalization ability of networks when sample size is small.

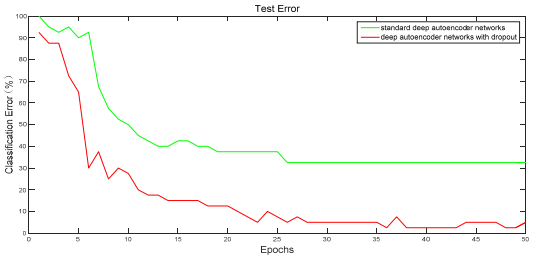

FIGURE VII. COMPARISON OF THE CLASSIFICATION ERROR OF STANDARD AND DROPOUT DEEP AUTOENDER NETWORKS.

IV. MODEL EVALUATION

In this section, face recognition based on deep autoencoder networks with dropout is evaluated. We try various dropout rates and the number of hidden units to observe generalization performance of the network.

We keep the network structure be 784-800-800-40 and train deep autoencoder networks with various dropout. Figure 8(A) shows classification error as a function of dropout rates. In particular, when dropout rate $p=0$, deep autoencoder networks with dropout equate to standard deep autoencoder networks. Compared with $p=0$, the classification error of the network decreases from 32.5% to 2.5% when $p=0.4$. In other words, the correct identification rate of the network in the small sample ORL database reaches 97.5%.

Figure 8(B) shows the number of hidden units have an effect on classification error. Maintaining $p=0.4$ fixed, we find deep autoencoder networks with dropout yield lower test error for any hidden units in the small sample ORL database.

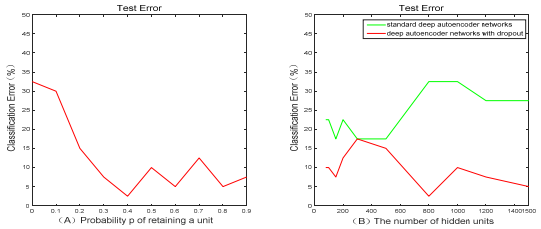

FIGURE VIII. THE CLASSIFICATION ERROR OF TEST DATA: (A)THE CLASSIFICATION ERROR AS A FUNCTION OF DROPOUT RATES. (B) THE CLASSIFICATION ERROR AS A FUNCTION OF THE NUMBER OF HIDDEN UNITS.

The key to effectively reduce over-fitting is that a deep autoencoder network with dropout randomly drop units from the neural network during training, which leads different networks at each training time. Sampling hidden units with a fixed probability p is equivalent to training a large number of deep networks with half hidden layer, that is, half nets. When we apply such networks to image recognition, each half nets can

A deep autoencoder network with dropout can effectively reduce over-fitting, since dropout randomly drop units from the neural network during training. This leads different networks at each training time. Sampling hidden units with a fixed probability p is equivalent to training a large number of deep networks with half hidden layer, each of which can bring a classification result when we apply such networks to face recognition. With training, the most networks can correctly recognize only few errors have little effect on whole result. What's more, we calculate actual output at test time in the sense of "mean network" at test time. All have contributed to improve the generalization ability and reduce over-fitting of deep autoencoder networks.

V. CONCLUSIONS

Deep autoencoder networks show excellent ability in learning feature, but its poor performance on test data go against visualization and classification of image. In order to improve the generalization ability and reduce over-fitting, we apply dropout to optimize the deep autoencoder networks. The nets take the visual features of facial image as input, and output the class information of image by *Logistic* function. We firstly obtain approximate optimal value by pre-training with autoencoder. Secondly, we establish contact with label information of the input image and yield optimal model of deep autoencoder networks by supervised fine-tuning. The test error is calculated when test images are applied to the model.

We pose deep autoencoder networks with dropout and achieve face recognition based on networks using the ORL database. Our experiments show that deep autoencoder networks with dropout yield significantly lower test error on the ORL dataset, and bring a new conception in pattern recognition with deep learning.

ACKNOWLEDGMENT

We would like to thank Speech Vision, Robotics Group and Cambridge University Engineering Department for warm-hearted help of The ORL Database of Faces download from their web site and excellent suggestions of face recognition. This work is supported by the National Natural Science Foundation of China (NSFC) 11301493.

REFERENCES

[1] GE Hinton, RR Salakhutdinov. Reducing the dimensionality of data with neural networks. Science, 2006, 313(5786):504-7

[2] JE Moody. Number of Parameters: An Analysis of Generalization and Regularization in Nonlinear Learning Systems[A]. NIPS 4[C]. San Mateo, CA, 1992.847-854

[3] SJ Nowlan, GE Hinton. Simplifying Neural Networks by Soft Weight-Sharing. Neural Computation, 1992, 4(4):473-493

[4] DJC Mackay. Bayesian Interpolation. Springer Netherlands, 1992, 4(3):415 - 447Springer Netherlands, 1992, 4(3):415 - 447

[5] S Amari, N Murata, KR Muller,M Finke, HH Yang. Asymptotic statistical theory of overtraining and cross-validation. IEEE Transactions on Neural Networks, 1997, 8(5):985-996

[6] CE Imrie, S Durucan. River Flow Prediction Using the Cascade-correlation Neural Network Learning Architecture. International Conference on Water 99: Joint Congress; Hydrology & Water Resources Symposium, 1999:94-99

[7] B Zhang, RS Govindaraju. Prediction of watershed runoff using Bayesian concepts and modular neural networks. Water Resources Research, 2000, 36(3):753–762

[8] M Yoon, Y Lee, S Lee, I Ivrissimtzis, HP Seidel. Surface and normal ensembles for surface reconstruction. Computer-Aided Design, 2007, 39(5):408-420

[9] R Salakhutdinov, A Mnih. Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. International Conference on Machine Learning, 2008:880-887

[10] HY Xiong, Y Barash, BJ Frey. Bayesian prediction of tissue-regulated splicing using RNA sequence and cellular context. Bioinformatics, 2011, 27(18):2554-2562

[11] GE Hinton, N Srivastava, A Krizhevsky,I Sutskever, RR Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. Computer Science, 2012, 3(4):págs. 212-223

[12] A Livnat, C Papadimitriou, J Dushoff, MW Feldman. A mixability theory for the role of sex in evolution. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105 (50):19803-8

[13] A Livnat, C Papadimitriou, N Pippenger, MW Feldman. Sex, mixability, and modularity. Proceedings of the National Academy of Sciences, 2010, 107(4):1452-7

[14] N Srivastava, G Hinton, A Krizhevsky, I Sutskever, R Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014, 15(1):1929-1958

[15] G Wang, N Shi, Y Shu, D Liu. Embedded Manifold-Based Kernel Fisher Discriminant Analysis for Face Recognition. Neural Processing Letters, 2016, 43(1):1-16

[16] S Gao, Y Zhang, K Jia, J Lu. Single Sample Face Recognition via Learning Deep Supervised Autoencoders. IEEE Transactions on Information Forensics & Security, 2015, 10(10):1-1

[17] Speech Vision, Robotics Group. Cambridge University Engineering Department. The ORL Database of Faces. http://am.ac.uk/research/dtg/attarchive/facedatabase.html