2019 2nd International Conference on Information Science and Electronic Technology (ISET 2019)

The Research of Deep Belief Network

Zizhao Chen[a], Wencheng Jiao

Department of Electronic and Optical Engineering Army Engineering University Shijiazhuang, China

[a]294782015@qq.com

Keywords: Deep Learning, DBN, RBM, input data, network structure, parameter optimization.

Abstract: As a hot research field in recent years, deep learning has great application prospects. First the technical research results and the application fields of deep learning is introduced in this paper. Second the DBN network model is taken as an example to introduce its theoretical basis, network structure and training process. Then we analyze the research thoughts and achievements of experts and scholars in recent years and propose to classify the main research directions into three categories, containing DBN input data, DBN network structure and DBN parameter optimization. At the same time, the literatures in various fields at home and abroad in recent years are classified according to the three categories mentioned. Finally, the research trends of the deep belief network are analyzed, in order to provide researchers with ideas for improving DBN.

1. Introduction

Nowadays, the research and application of deep learning technology is in the ascendant. In terms of technical research, network models are proposed such as Automatic Encoder (AE), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Deep Belief Network (DBN). In the application of technology, many popular areas are promoted such as Face Recognition, Self-driving Car and Big Data Processing. This paper takes the deep belief network as an example to introduce its basic theory and research results in recent years.

At the beginning of the development of deep learning, the traditional neural network [1] has been unable to achieve the needs of people for fast extraction and accurate learning. In 2006, Hinton [2] proposed a fast learning algorithm applied to DBN networks, opening a new door to combine deep learning with production and life. At present, DBN network technology is mainly used in the fields of text detection, face and expression recognition, remote sensing image classification, fault diagnosis and so on [3], but there is a lack of relevant literature in technical research.

This paper summarizes the research ideas and achievements of scholars in various fields on DBN networks, and divides DBN research into three aspects: input data, network structure and parameter optimization. In the first chapter, the research background and problems of the deep belief network is introduced. In the second chapter, the basic principles and training process of the deep belief network is described. In the third chapter, the research classification and viewpoints of the deep belief network is discussed. In the fourth chapter, we summarize the deep belief network development and look to the future.

2. Model of Deep Belief Network

The Deep Belief Network (DBN) is a probability generation model. The generation model is to establish a joint distribution between observation data and labels, and infer the distribution of data samples by joint distribution. The traditional DBN uses the Gibbs sampling algorithm to sample the unknown probability density, the contrast divergence (CD-k) algorithm to optimize the sampling rate, the greedy algorithm to train the hidden layer and the BP algorithm to reverse global tuning. The neural network generates training data according to the principle of maximum probability, forms high-level abstract features, and learns and classifies training data according to abstract features.

2.1. Restricted Boltzmann Machine (RBM) Model

DBN is a type of neural network consisting of an input layer (visual layer), a hidden layer, and an output layer. The difference between DBN and other neural networks is that DBN is composed of multi-layer Restricted Boltzmann Machines (RBMs), and the network structure is relatively simple.

The RBM (shown in Figure 1) is based on the Boltzmann machine (BM) and the conditions in which the same layer of neurons cannot be connected to each other. Its essence is an undirected graph model composed of random visible neurons and hidden neurons. The undirected graph model guarantees conditional independence when neurons are disconnected, that is, nodes depend only on nodes adjacent to them, and conditions are independent of any other nodes that are not connected to them. As shown in the figure, the Gibbs sampling algorithm repeatedly iteratively adjusts the weights between neurons to obtain a joint probability distribution closest to the training samples, so that the hidden layer can extract data features more abstractly and accurately. The CD-k algorithm is an algorithm proposed by Hinton to accelerate the optimization of Gibbs sampling. Based on the one-step contrast divergence algorithm (CD-1), the RBM training process is as follows:

Step 1: Input the training data X (indicated by v) and initialize the RBM network weights;

Step 2: Calculate the activation probability of each hidden layer neuron according to (1):

$$P(h_j^0 \mid v^0) = \sigma(W_j v^0 + a_j) \tag{1}$$

Step 3: Activate the hidden layer and determine the activation probability of each hidden layer neuron according to (2):

$$\begin{aligned} P(h_j^0) &= 1 \quad if \quad P(h_j^0 \mid v^0) > rand(0,1) \\ P(h_j^0) &= 0 \quad if \quad P(h_j^0 \mid v^0) < rand(0,1) \end{aligned} \tag{2}$$

Step 4: Reconstruct the display layer and calculate the activation probability of each layer of neurons according to (3):

$$\begin{aligned} &P(v_i^1 \mid h^0) = \sigma(W_i h^0 + b_i) \\ &P(v_i^1) = 1 \quad if \quad P(v_i^1 \mid h^0) > rand(0,1) \\ &P(v_i^1) = 0 \quad if \quad P(v_i^1 \mid h^0) < rand(0,1) \end{aligned} \tag{3}$$

Step 5: Calculate and update the weight according to (4):

$$\begin{aligned} &P(h_j^1 \mid v^1) = \sigma(W_j v^1 + a_j) \\ &W_{ij} = W_{ij} + \lambda(P(h^1 \mid v^1)P(v^1) - P(h^0)P(v^0)) \end{aligned} \tag{4}$$

Where i is the number of nodes in the layer of neurons; j is the number of nodes in the hidden layer; σ is the activation function, usually is the sigmoid function; λ is the weight learning rate.

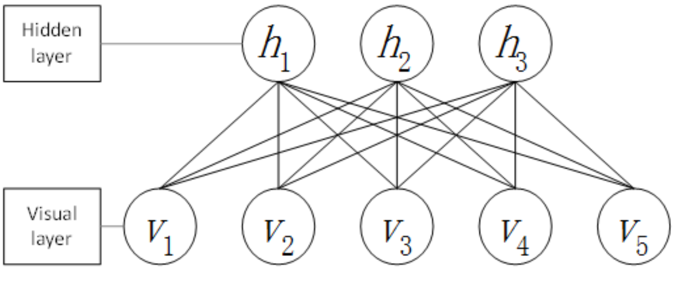

Fig. 1 Structural Model of RBM

2.2. Deep Belief Network (DBN) Model

This paper takes a three-layer restricted Boltzmann machine as an example (shown in Figure 2), and introduces its training steps in combination with the greedy learning algorithm and the BP back propagation algorithm:

Step 1: Repeat the iteration N times according to the RBM training steps introduced in the previous section to improve the accuracy of the network parameters.

Step 2: According to the greedy learning algorithm, the i-th RBM output data is used as the i+1th layer input data, and the DBN model is trained layer by layer;

Step 3: According to the BP back propagation algorithm, reverse optimization of each layer to search for the global (local) best advantage;

Step 4: Repeat step 3 until the maximum number of network searches is reached or the global (local) best advantage is found;

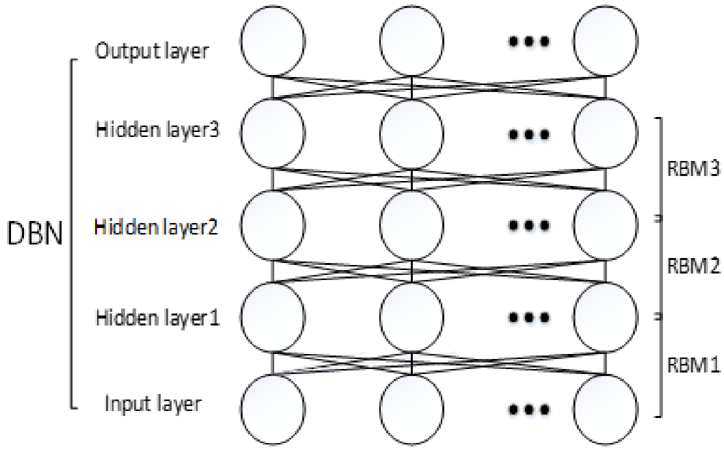

Fig. 2 Structural Model of DBN (3 hidden layer)

3. Classification of deep BELIEF network

In practical applications, an excellent network model can often make training, learning and recognition more effective. This chapter mainly analyzes and summarizes the application on DBN in various fields, and proposes to divide DBN research into DBN input data aspect, DBN network structure aspect and DBN parameter optimization aspect.

3.1. Input data of DBN

DBN input data research is a pre-processing study of network input data, which mainly includes two small pieces of input data type research and input data processing research.

In the research of input data types, the traditional DBN can only be applied to binary discrete data. The article [4] points out and proves that the visible variable distribution and the hidden variable distribution can be any exponential family distribution. Extending the RBM binary discrete input to any exponential family of distributed inputs makes traditional DBNs more widely applicable, but binarizing information inevitably loses valid information. In reference [5], the input is binary state, and the RBM variant models such as Gaussian RBM and mean covariance RBM model are discussed. The selection steps of Gaussian RBM and mean covariance RBM Hamiltonian annealing importance sampling model is deduced in detail. While using the ISOLET data set to verify the correctness of the model. When the nonlinear neural network is doing nonlinear transformation, the activation probability will gradually shift as the depth of the network deepens, which makes the convergence process slower. Article [6] adds batch normalization (BN) to standardize the input data in the DBN network. By normalizing the data before each layer of the input excitation function, the input distribution is always in an unsaturated state. The article proves that IDBN is more advantageous through experiments.

In the research of input data processing, it mainly deals with training data to expand the scope of application of the network and improve the running speed of the network. In reference [7], the multi-dimensional time series is used as input, and the Principal Components Analysis

(PCA) is used to reduce the dimension in the sample data dimension, compress the sample data volume, and then use the Dynamic Time Warping (DTW) to calculate the distance matrix of different data. The obtained matrix is used to generate a one-dimensional vector and then input into the DBN for training. The assumption for this scenario is that the sample input is not missing data but is too harsh for actual production. In reference [8], the spline function is established by the cubic spline difference algorithm, and the data points of the vacancy samples are replaced by the nearest new data points, thereby solving the problem that the input vector dimensions are not uniform and the data is missing. The papers [9-11] use wavelet transform (WT), wavelet packet transform (WPT), empirical mode decomposition (EMD), and protein quality assessment (QA) to process the acquired signals and input them into the DBN network for learning. A variety of data processing methods have broadened the scope of application of the network to some extent.

Aiming at the problem of RBM efficiency, the article [12] proposes the concept of generalized redundancy elimination without affecting the classification accuracy. Two techniques, bounds-based filtering and delta product, are extended to optimize a large number of repeated calculations. At the same time, the GPU's parallel architecture is used to effectively combine aggregated warp filtering and dual copies for coalescing. The results show that the generalized elimination model is faster than normal, with 2 to 3 times improvement in training and 5 times in prediction. In reference [13], a PA_DBN parallel acceleration strategy is proposed to improve efficiency, including using partition algorithm to solve data skew and introducing resilient distributed data (RDD) caching algorithm to solve data reuse. Experiments show that PA_DBN is faster than traditional DBN, and it also verifies the combination of four feature indexes of NDVI, DVI, RVI and EVI and the impact of different network structures on accuracy. The article [14] solves the problem of parameter input initialization of RBM network by adding sparse noise reduction self-encoder. At the same time, the SDAE network is used to pre-train the input data, and then the trained network weights are assigned to the RBM initialization, thereby speeding up the network operation.

The introduction of papers on input data is shown in Table 1. In other respects, the article [15] uses Bagging method to generate different training subsets, and then uses ELM extreme learning machine to train individual classifiers to obtain different sub-set data as input data for training. The article [16] will integrate learning and Bagging algorithm Random sampling training yields the input projection vector and then enters the DBN network for training. Both methods make the network superior in applications by processing the input data. In reference [17], the convolutional neural network is combined with the deep belief network, and the CNN is used to convolve and reduce the graph, and then the hidden layer feature extracted from CNN full connection layer is used as the input data of DBN. Experiments show that this method is superior to the existing scheme. The papers [18-20] use feature fusion technology to fuse different attribute characteristics of input data, thereby improving the accuracy of network identification and classification.

Table.1. INPUT DATA of DBN

Authors	method	Advantages	shortcoming
Welling et al.[4]	Gauss RBM;	Wider applicability	No other nonlinear distribution is involved
Peng Li-xia[5]	Two RBM models;	Gives the method and steps for selecting two different RBMs	Only discuss model selections under a single dataset
Li H, Zhang ZP[6]	batch normalization	Increase gradient descent, convergence process speed	The accuracy rate is only about 85%, there is still room for improvement.
Ge Qiangqiang[7]	CPCA;DTW;	Greatly compress data using CPCA and DTW	The program assumes that the conditions are too harsh and not suitable for actual production.
Shan Jiusi[8]	Spline difference; RBF;	Solve the input vector dimension is not uniform	Brought precision reduction and increased training time
Huang H B et al.[9]	Continuous RBMs	Learning voice signals using continuous restricted RBM	No comparison with other RBM or DBN models

H.Z Wang et al.[10]	WT;QR;	Processing data characteristics using WT method and QR method	Reliability has declined
Cao, R et al.[11]	Protein quality assessment;	Select features with theoretical support and representativeness	Only for a single field
Ning L et al.[12]	Redundancy Elimination	Provides an optimization idea	Did not give sufficient experimental proof
Ying C et al.[13]	Partition algorithm; Feature index	Optimize data input speed with computer index technology	Requires processing image data first
Zeng An et al.[14]	Sparse noise reduction self-encoder	Improve local optimization problems and improve accuracy	The experiment is not enough, only consider the double hidden layer and the number of fixed nodes
Yu L et al.[15]	Integrated learning;ELM;	Train individual classifiers with ELM to get different subsets	Predetermined architecture of the DBN model
Z Zhao et al.[16]	SAR;Integrated learning;Bagging;	Random sampling using the Bagging algorithm to optimize data	Image processing is more complicated
Tang B et al.[17]	CNN;	Dimensional optimization of input data using CNN network	The accuracy of each model in the comparison experiment is generally low.
Hu Yongtao[18]	FCM;PCA;Multi-feature fusion	Feature fusion using feature fusion method	Adoption method is more traditional, no strong innovation
Xu Yongqing et al.[19]	Multi-feature fusion	Multi-feature fusion extraction feature scheme	The accuracy of the calculation model method is only 85%.
Li Shuang[20]	Multi-feature fusion;Image Processing;	Fusion of different features of the image as input data	Added feature extraction and training time

3.2. Network structure of DBN

In 2009, Professor Bengin [21] proposed an open question, how to determine the appropriate network depth? It has led many scholars to study the DBN network structure. The research of DBN network structure mainly focuses on the research of DBN hidden layer number, the number of neurons in each layer and the improvement of DBN network structure.

For the problem that the number of DBN hidden layers and the number of nodes cannot be determined, the article [22] provides an important idea for the identification of hidden layers: each layer of the hidden layer is to gradually reduce the samples correlations, that is, to reduce the common features existing between samples. The article draws on the theory of signal detection and estimation, and solves the mutual correlation between the h values of each layer, and uses the average mutual relationship as a basis for measuring whether the hidden layer is reasonable. The article [23] combined with information entropy to determine the number of hidden layer neurons. The ratio of the information entropy of the hidden layer to the information entropy of the input layer is equivalent to the ratio of the number of neurons in the hidden layer to the number of neurons in the input layer, thereby determining the number of neurons in the hidden layer.

Table.2. Network Structure of DBN

Authors	method	Advantages	shortcoming
Gao Qiang, Ma Meimei[22]	Correlation coefficient; detection and valuation;	Measure hidden layers by average cross-correlation	Longer training time under the given optimal conditions
Liao Qiang, Zhang Jie[23]	Information entropy;	Determine the number of neurons by hidden layer and input layer information entropy	Need to artificially set the reconstruction error
Pan Guangyuan et al.[24]	Reconstruction error;Energy function;	Compare the reconstruction error with empirical knowledge to determine the number of layers	Ignore the relationship between different hidden layers
Guo C et al[25].	Integrated classification	Ability to choose the optimal number of hidden layers and the number of nodes	Method is time consuming and laborious
Qiao J et al.[26]	Self-organizing neural network; root mean square error;	Dynamically determining the number of neurons by combining SODBN and RMSE	Failed to use accuracy as a parameter of choice
Zhang Shizhen et al.[27]	Dynamic increase and decrease branching algorithm;	Determine the number of neurons by dynamic increase and decrease branch algorithm	Too many parameter settings, complex algorithms, lack of robustness
Zhang Yikang[28]	SOM; glial cells;	Add glial cell reinforcement learning to the hidden layer	Increase the difficulty of searching for optimal parameters
GAO Qiang[29]	Optimal receiver	Equivalent DBN network as the best receiver model	Only give theoretical proof, fail to prove experimentally
Ma Meimei[30]	Optimal receiver;Best decision threshold;		Experimental proof of lack of large data sets
Wang D et al.[31]	CNN;	Combine CNN network with DBN network	Increased network complexity
Siyuan W	Time series; nonlinear prediction	Add parameters associated with past times	Network robustness is weakened

The article [24] adopts an integrated classification method, and accesses SVM, KNN and RBF three classifiers after each layer of hidden layer. By comparing the output accuracy of the three classifiers, the number of hidden layers and the number of nodes selected by the DBN network are determined. But this method is time consuming and laborious, but it is also a way at the moment. In reference [25], a DBN network depth determination method based on reconstruction error is proposed. According to the initial value of subjective design, the number of network layers is adaptively increased. The article [26] proposes a self-adjustable network model SODBN. The SODBN model compares the root mean square error (RMSE) of the current time with the previous time by setting the initial threshold, and compares the peak intensity of the single neuron with the set peak intensity threshold to determine the increase or decrease of the hidden layer number and the number of nodes. However, the program inevitably introduces artificial factors. The model proposed in [27] optimizes the DBN model by using an algorithm that dynamically increases the number of hidden layers, the number of hidden layer neurons, and dynamically reduces the number of hidden layer neurons. At the same time, the theory proves that the algorithm has convergence and practicability. The experimental results show that the DBN network using the dynamic increase and decrease branching algorithm has a large performance improvement.

Aiming at the problem of DBN network structure, Scheme [28] found the mechanism that the glial cell stimulates the activation of hidden neurons, thereby adding a stimulating layer of glial cells to enhance the activation probability of hidden neurons in the DBN network. The improved scheme has a significant improvement in convergence speed and accuracy rate compared to the conventional

DBN. However, while increasing the hidden layer structure, it increases the difficulty of neural network weight initialization and finding the global best. In reference [29], from the "best receiver" equivalent model theory and the multiple data set comparison experiments, the DBN network with weighted interval is proved to have better performance. On the basis of the above paper, the article [30] is proved by formulas that the process of determining the weight in the RBM network model is the process of determining the input signal by the best receiver. After the original signal passes through the noise channel, it enters the optimal decision machine, and the best decision machine decides to restore the original transmission signal according to the probability. And the article further proves that the DBN model of multi-hidden and multi-neurons can be equivalent to the best receiver with N-ary input signal.

The fusion of different neural networks can achieve the complementarity of superiority and inferiority. In reference [31], by adding the convolutional layer and the pooling layer in the DBN network, the learning of edge probability and feature correlation is enhanced, which makes the face feature recognition with obscuration more superior than traditional DBN and CNN. The accuracy rate is as high as 98.9%. The paper [32] proposed a nonlinear prediction model CRBM for time series prediction. The CRBM model correlates the parameters of the past time display layer or hidden layer with the parameters of the current time display layer or hidden layer, so as to predict the time series. The article further proposes to add two parameters related to the past moments based on its two-layer CRBM, which makes the model prediction more accurate. For RBM, only first-order Markov sequences can be learned from the data, and high-order time delays or loops cannot be expressed in the form of feedback. Combined with a recurrent neural network (RNN), the scheme [33] implements a variable order Boltzmann (VBM) by adding a dynamic Gaussian Bayesian fitness function. The VBN is composed of VBMs and successfully applied to text prediction. In the context of reinforcement learning (RL), the article [34] proposes a CRBM structure and related training programs. The model has the ability to learn more complex and higher-dimensional information by including an active node in the hidden layer that retains the state and operation of the previous sub-policy. The introduction of papers on network structure is shown in Table 2.

3.3. Parameter optimization of DBN

In 2010, Professor Hinton [35] based on years of research experience and gave reasonable advice on DBN parameter settings. The research on DBN parameter optimization mainly focuses on the research of over-fitting problem, the research of learning rate and the research of reverse search algorithm.

In response to the over-fitting problem of deep learning, Professor Hinton [36] proposed a coping strategy for dropout. By randomly discarding certain neurons with probability p during the training process, the input data is learned in a new network structure, and then the network weights after the training is p-scaled as the final network weight. The experimental results show that the DBN network with dropout sparse performance is better. In reference [37], a sparse penalty factor is added to the traditional Gaussian RBM to optimize the cost function. The article sets the conditional expectation, trade-off parameters and sparse constants of the data to control the sparsity, which makes the network parameter adjustment more complicated. The literature optimizes the sparse regularization term based on Laplace function [38], error square sum [39] and rate distortion theory [40] to form a sparse deep confidence network. The improved scheme suppresses the over-fitting phenomenon to a certain extent and improves the accuracy.

The traditional DBN network learning speed is fixed. If it is too large, it will cause the convergence process to fluctuate drastically. If it is too small, the convergence speed will be too slow. In reference [41], an adaptive learning rate method is proposed, which makes the DBN learning rate self-adjust according to the direction of two reconstruction errors. When the direction of the two reconstruction errors are the same, the learning rate is increased. When the direction of the error direction is different, the learning rate is reduced, thereby speeding up the network convergence. The article [42] introduces the adaptive step size concept and propose the ADBN algorithm. The independent step size self-adjusting parameter algorithm is used instead of the global learning rate in the original DBN network. Experiments show that on the MNIST dataset, the ADBN algorithm significantly speeds up the training compared to the traditional DBN algorithm. According to the DBN_BP algorithm, the learning rate in the actual process is changed from large to small, and the change range of weight tends to be stable. The paper [43] proposes a method of alternately updating the number of layers. In the first few cycles, the original loss function is used to update the formula, and when the learning rate becomes small, the way to update the highest layer of the network or the highest layer of the network is used.

Aiming at the problems of long gradient search time and slow efficiency of traditional gradient descent algorithm (GD), the paper proposes to use the heuristic search algorithm such as harmony algorithm [44], particle swarm algorithm [45] and genetic algorithm [46,47] to optimize the reverse trimming stage, speed up search efficiency and improve the accuracy of classification. The article [48] abandons the traditional single hidden layer Gibbs sampling, and discusses the combination of four multi-layer Gibbs sampling: two by two non-nesting, two by two nesting, incremental non-nesting, and incremental nesting. Taking the four-layer and three-layer hidden layers as examples, it is concluded that the multi-hidden layer Gibbs sampling effect is better than the traditional DBN network model. Combining the advantages of traditional one-step Gibbs sampling and step-by-step Gibbs sampling, the article [49] proposes a dynamic Gibbs sampling that combines the two. Set m1 and m2 as thresholds. When the number of iterations is less than m1, one-step Gibbs sampling is used. When the number of iterations is greater than m1 and less than m2, Gibbs sampling with varying steps is used. When the number of iterations is greater than m2, Gibbs sampling with fixed steps is used.

In other respects, article [50] combines deep confidence networks with fractional calculus. It uses the form of fractional calculus to update the weight and network parameters such as offset. The weight (offset) at the next moment is calculated by taking the weight (offset) of the previous moment, the weight (offset) of the current time, and the weight (offset) change of the current time as parameters. The experimental results show that on the Berlin and Telugu databases, the network accuracy rates are 98.39% and 95.88%, respectively. In reference [51], the fuzzy set is used to replace the common set parameters such as weights and offsets in the model, and the energy function is fuzzified, and then the probability value is determined. Experiments show that FRBM has better robustness than RBM. The introduction of papers on parameter optimization is shown in.

Table.3. Parameter optimization of DBN

Authors	method	Advantages	shortcoming
Hu G et al.[36]	Drop-outStrategy	Dynamic random discarding of neurons to deal with over-fitting problems	Reduced network accuracy and stability
Liu Mengxi et al.[37]	Sparse punishment; GRBM	Add sparse penalty factor to optimize cost function	Adding artificial parameters to increase network complexity
Xu Yi et al.[38]	Laplace function	Using Laplacian functions to sparse regularization	
Ekanadham C[39]	Square sum error;	Sparse regularization with the sum of squared error functions	
Ji N N et al.[40]	Rate distortion theory;	Relative entropy function as loss function	
Qiao Junfeiet al.[41]	Adaptive learning rate	Automatically adjust the learning rate according to the reconstruction error	Set the number of network nodes and hidden layers by experience
Yang Chunde, Zhang Lei[42]	Adaptive step size;	Replace the global fixed learning rate with a self-adjusting step	Replace the global fixed learning rate with a self-adjusting step
Li Xuan, Li Chunsheng[43]	Alternate transformation; learning rate	Alternately transform update weights, increase training speed	Alternate training may result in incomplete training of high-level hidden layers
João[44]	Harmonic algorithm	Optimize search speed by	The high load calculation of

Kuremoto[45]	Particle swarm algorithm	combining different search algorithms with gradient descent algorithm	the improved algorithm requires higher performance on the computer.
Levy E et al.[46]	Genetic algorithm		
Liu K et al.[47]	Genetic algorithm		
Li Fei et al.[48]	Multi-layer Gibbs sampling;	Try a combination of four multi-layer Gibbs sampling	As the number of Gibbs sampling steps increases, the training time will increase dramatically.
Shi Ke et al.[49]	Dynamic Gibbs sampling;	Dynamically determine the number of Gibbs sampling steps	

4. Summary and outlook

With the increasing amount of data and the expansion of application areas, the challenges faced by DBN networks are gradually increasing. On the one hand, it is necessary to step up computer GPU and clustering research to improve hardware support [52]. On the other hand, optimizing the computational efficiency and structural simplicity of neural networks to enhance software support is also essential. Based on the background of deep learning, this paper summarizes the research results of experts and scholars in deep belief network (DBN) technology in recent years. In the future, researchers can continue to optimize DBN performance in the following ways:

(1) Combine computer index technology to optimize input data speed and pre-processing input data;

(2) Combine other neural networks to build a novel network structure and explore how to dynamically determine the network structure;

(3) The problem of under-fitting and over-fitting in DBN networks is a hot topic at present. How to solve problems more subtly is one of the goals of scholar.

Acknowledgements

Electronic information equipment remote maintenance support technology(NO:41400034102).

References

[1] Kohonen T An introduction to neural computing[J]. Neural Networks, 1988, 1 (1):3-16.

[2] Hinton G E, Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets [J]. Neural Computation, 2014, 18 (7):1527-1554.

[3] Liu Fang-yuan, Wang Shuihua, Zhang Yidong. Overview of Deep Trust Network Model and Application Research [J]. Computer Engineering and Applications, 2018, 54(01): 11-18+47.

[4] Welling M, Rosen-Zvi M, Hinton G. Exponential family harmoniums with an application to information retrieval [C]// International Conference on Neural Information Processing Systems. MIT Press, 2004.

[5] Peng Li-xia. Research on model selection problem of deep belief network [D]. Southwest Jiaotong University, 2018.

[6] Li Hui, Zhang Zhi-pan. Research on Fault Diagnosis of Power Transformer Based on Improved Deep Belief Network [J]. Software Guide.

[7] 2018,17 (07):169-172.

[8] Ge Qiang-qiang. Research on Data Driven Fault Diagnosis Method Based on Deep Confidence Network [D]. Harbin Institute of Technology, 2016.

[9] Shan Jiu-si. Research and implementation of railway turnout fault diagnosis system based on deep confidence network [D]. Shijiazhuang Railway University, 2017.

[10] Huang H B, Li R X, Yang M L, et al. Evaluation of vehicle interior sound quality using a continuous restricted Boltzmann machine-based DBN [J]. Mechanical Systems and Signal Processing, 2017, 84:245-267.

[11] H.Z. Wang,G.B. Wang,G.Q. Li,et al. Deep belief network based deterministic and probabilistic wind speed forecasting approach [J]. Applied Energy,2016,182.

[12] Cao, R., Bhattacharya, D., Hou, J. et al. DeepQA: improving the estimation of single protein model quality with deep belief networks [J]. BMC Bioinformatics (2016) 17: 495.

[13] Ning L, Pittman R, Shen X. POSTER: Cutting the Fat: Speeding Up RBM for Fast Deep Learning Through Generalized Redundancy Elimination [C]// International Conference on Parallel Architectures & Compilation Techniques. IEEE Computer Society, 2017.

[14] Ying, C., Huang, Z. & Ying, C. Accelerating the image processing by the optimization strategy for deep learning algorithm DBN[J]. Wireless Com Network (2018) 2018: 232.

[15] Zeng An, Zhang Yi-nan, Pan Dan, et al. Deep confidence network based on sparse noise reduction self-encoder [J]. Journal of Computer Applications, 2017, 37 (09): 2585-2589.

[16] Yu L, Yang Z , Tang L . A novel multistage deep belief network based extreme learning machine ensemble learning paradigm for credit risk assessment [J]. Flexible Services & Manufacturing Journal, 2016, 28 (4):576-592.

[17] Zhao Zhi-qiang ,Jiao Li-cheng, Zhao Jia-qi. Discriminant deep belief network for high-resolution SAR image classification [J]. Pattern Recognition, 2017, 61:686-701.

[18] Tang B , Liu X , Lei J , et al. DeepChart: Combining deep convolutional networks and deep belief networks in chart classification [J]. Signal Processing, 2016, 124:156-161.

[19] Hu Yong-tao. Bearing fault diagnosis based on multi-feature fusion and deep belief network [D]. Yanshan University, 2017.

[20] Xu Qing-yong, Jiang Shun-liang, Huang Wei,et al. Image Classification Algorithm for Deep Belief Network Based on Multi-feature Fusion[J]. Computer Engineering, 2015, 41(11): 245-252.

[21] Li Shuang. Image Classification Algorithm Based on Multi-feature Fusion and Deep Learning [J]. Journal of Henan University of Science and Technology (Natural Science Edition), 2018, 46(04): 50-55.

[22] Bengio, Y. Learning Deep Architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2 (1):1-127.

[23] Gao Qiang,Ma Yan-mei.Research and Application of the Number of Layers of Deep Belief Network (DBN) Network[J].Science Technology and Engineering,2016,16 (23):234-238+262.

[24] LIAO Qiang,ZHANG Jie.Optimization of DBN Network Structure Based on Information Entropy[J].information communication,2018 (01):44-48.

[25] Pan Guang-yuan, Chai Wei, Qiao Junfei. Depth determination method of DBN network[J]. Control and Decision, 2015, 30 (02): 256-260.

[26] Guo C , Yang Y , Pan H , et al. Fault Analysis of High Speed Train with DBN Hierarchical Ensemble[C]// International Joint Conference on Neural Networks. IEEE, 2016.

[27] Qiao J , Wang G , Li X , et al. A self-organizing deep belief network for nonlinear system modeling[J]. Applied Soft Computing, 2018:S1568494618300255.

[28] Zhang Shi-zhen, Song Wei, Wang Chen-ni,et al. DBN model for optimizing network structure using dynamic increase and decrease algorithm [J/OL]. Computer Science and Exploration: 1-14[2018-11-26].

[29] Zhang Yi-kang. Research on deep learning method based on restricted Boltzmann machine[D]. Beijing University of Chemical Technology, 2016.

[30] GAO Qiang,WANG Ming.Research on Equivalent Model and Weight Extension Algorithm of Deep Belief Network[J].Electrical Measurement & Instrumentation,2017,54(23):54-59.

[31] Ma Yan-mei. Research and application of deep belief network (DBN) equivalent model [D]. North China Electric Power University, 2017.

[32] Wang D , Li M , Li X . Face Detection Algorithm Based on Convolutional Pooling Deep Belief Network[C]// International Conference on Electrical. 2017.

[33] Siyuan W ; Yong L ; Xu Z. A differentiated DBN model based on CRBM for time series forecasting[C]. 2017 IEEE 17th International Conference on Communication Technology (ICCT),Chengdu China,Oct 2017.IEEE,2018:1926-1931.

[34] Chaturvedi I , Ong Y S , Tsang I W , et al. Learning word dependencies in text by means of a deep recurrent belief network[J]. Knowledge-Based Systems, 2016:S0950705116302349.

[35] Djurdjevic P D , Huber M . Deep Belief Network for Modeling Hierarchical Reinforcement Learning Policies[J]. 2013.

[36] Hinton G E . A practical guide to training restricted Boltzmann machines. [J]. Momentum, 2012, 9(1):599-619.

[37] Hu G, Li H, Luo l, et al. An improved dropout method and its application into DBN-based handwriting recognition[C]. IEEE 2017 36th Chinese Control Conference(CCC), 2017.

[38] Liu Meng-xi, Wang Zheng, Song Jiu-xu, et al. Research on Image Classification and Recognition Based on Sparse Depth Confidence Network[J].Microelectronics&Computer,2018,35 (09):59-63.

[39] Xu Yi, Li Wei, Song Wei. Research on improved deep confidence network classification algorithm [J/OL]. Computer Science and Exploration: 1-15[2018-11-26].

[40] Ekanadham C . Sparse deep belief net model for visual area V2[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2007.

[41] Ji N N , Zhang J S , Zhang C X . A sparse-response deep belief network based on rate distortion theory[J]. PATTERN RECOGNITION.

[42] Qiao Jun-fei, Wang Gong-ming, Li Xiao-li, et al.Design and application of deep belief network based on adaptive learning rate[J].Acta Automatica Sinica,2017,43(08):1339-1349.

[43] YANG Chun-de, ZHANG Lei. Image classification method based on adaptive deep confidence network[J]. Computer Engineering and Design, 2015, 36(10): 2832-2837.

[44] LI Xuan, LI Chun-sheng. A fast training method for DBN-DNN with alternating transform layer number[J].Application Research of Computers,2016,33(03):843-847.

[45] João Paulo Papa, Scheirer W , Cox D D . Fine-tuning Deep Belief Networks using Harmony Search[J]. Applied Soft Computing, 2016, 46(C):875-885.

[46] Kuremoto T , Kimura S , Kobayashi K , et al. Time Series Forecasting Using Restricted Boltzmann Machine[J]. Communications in Computer & Information Science, 2012, 304:17-22.

[47] Levy E , David O E , Netanyahu N S . [ACM Press the 2014 conference - Vancouver, BC, Canada (2014.07.12-2014.07.16)] Proceedings of the 2014 conference on Genetic and evolutionary computation - GECCO \"14 - Genetic algorithms and deep learning for automatic painter classification[J]. 2014:1143-1150.

[48] Liu K , Zhang L M , Sun Y W . Deep Boltzmann Machines Aided Design Based on Genetic Algorithms[J]. Applied Mechanics and Materials, 2014, 568-570:4.

[49] LI Fei, GAO Xiao-guang, WAN Kai-fang. Research on RBM Training Algorithm Based on Dynamic Gibbs Sampling[J].Acta Automatica Sinica,2016,42(06):931-942.

[50] Shi Ke, Lu Yang, Liu Guang-liang, Bi Xiang, et al. Deep belief network training method based on multi-hidden layer Gibbs sampling [J/OL]. Acta Automatica Sinica: 1-10[2018-11-26].

[51] Mannepalli K , Sastry P N , Suman M . FDBN: Design and development of Fractional Deep Belief Networks for speaker emotion recognition[J]. International Journal of Speech Technology, 2016, 19(4):1-12.

[52] Chen C L P, Zhang C Y , Chen L , et al. Fuzzy Restricted Boltzmann Machine for the Enhancement of Deep Learning[J]. IEEE Transactions on Fuzzy Systems, 2015, 23(6):2163-2173.

[53] Chandra B, Sharma R K . Fast learning in Deep Neural Networks[M]. Elsevier Science Publish B. V. 201.