www.s2is.org
International Journal on Smart Sensing and Intelligent Systems
ISSN 1178-5608

HYPERSPECTRAL DATA FEATURE EXTRACTION USING DEEP BELIEF NETWORK

Jiang Xinhua[1], Xue Heru[*,1] , Zhang Lina[2] and Zhou Yanqing[1]

1. College of Computer and Information Engineering,
Inner Mongolia Agricultural University, 010018
Hohhot, Inner Mongolia, P.R. China
Emails: jiang-xh@163.com

2. College of Physics and Electronic Information,
Inner Mongolia Normal University, 010022
Hohhot, Inner Mongolia, P.R. China

Submitted: July 15, 2016 ***Accepted: Oct. 26, 2016*** ***Published: Dec. 1, 2016***

Abstract- Hyperspectral data has rich spectrum information, strong correlation between bands and high data redundancy. Feature band extraction of hyperspectral data is a prerequisite and an important basis for the subsequent study of classification and target recognition. Deep belief network is a kind of deep learning model, the paper proposed a deep belief network to realize the characteristics band extraction of hyperspectral data, and use the advantages of unsupervised and supervised learning of deep belief network, and to extract feature bands of spectral data from low level to high-level gradually. The extracted feature band has a stronger discriminant performance, so that it can better to classify hyperspectral data. Finally, the AVIRIS data is used to extract the feature band, and the SVM classifier is used to classify the data, which verifies the effectiveness of the method.

Index terms: Hyperspectral, Feature extraction, Deep learning, Deep belief network, Restricted boltzman machines

I. INTRODUCTION

Hyperspectral image consists of a number of narrow and continuous spectrum, and it includes a wealth of detailed spectral characteristics of the objects, and can characterize features of the object by the spectral properties and geometry aspects. Hyperspectral data is composed by a large number of high-dimensional spectral data, has good correlation between bands and high information redundancy, and the spectral data has non-linear characteristics, the spectral analysis process vulnerable to "Hughes" phenomena influence [1,2], so a good performance multi-spectral analysis algorithm becomes impractical or low efficiency in high spectral analysis. In order to take full advantage of high-dimensional spectral characteristics as well as to reduce the data redundancy caused by complex calculations, we must reduce the hyperspectral image data dimensionality, and find more effective low-dimensional features to express the original high-dimensional data, meanwhile improve spectral image classification recognition accuracy.

Scholars for hyperspectral image data dimensionality reduction algorithm has been a lot of research, and achieved certain results. These methods are mainly divided into feature selection and feature extraction two kinds methods, the feature selection method is to select several band for subsequent processing by the algorithm directly from the original band, the feature extraction method is to transform a high-dimensional spectral space to a low-dimensional feature space through an algorithm, and maintain the original characteristics of spectral data. Classic hyperspectral dimensionality reduction methods are Principal Component Analysis (PCA) [3] and Linear Discriminant Analysis (LDA) [4]. Studies suggest that the linear PCA dimension reduction method abandoned property details of the spectral dimension, the result of dimensionality reduction of hyperspectral is same to multi-spectral image, and lost the original hyperspectral information, is not optimal method[3,5-6]. Thus, in recent years the research with nonlinear dimensionality reduction algorithms have become focuses, and include transformation based on kernel method and Manifold Learning algorithm. The kernel principal component analysis (Kernel PCA, KPCA) [7] and Kernel Discriminant Analysis (KLDA) [8] are representative kernel method. The manifold learning algorithms applied to hyperspectral image dimensionality reduction mainly include local linear embedding (LLE), isometric mapping (Isomap), Laplacian Eigenmaps (LE), local tangent space alignment(LTSA) [9-12] and so on. However, manifold learning algorithms are unsupervised dimensionality reduction methods, and they are difficult to improve the results of following analysis naturally.

The deep structure learning model can extract the characteristic data of the expressive power from the original data through the design of the deep structure, which has attracted more and more attention of scholars both at home and abroad in recent years in recent years. Deep belief network (DBN) is a kind of deep learning model, it proposed by Hinton in 2006 [13], the DBN model simulate the organization structure of the depth brain, and learn abstraction characteristics at different levels through a bottom-up automatic method, and obtain the nonlinear feature description of analyzing objects. DBN combines the advantages of unsupervised learning and supervised learning, and it is an automatic extraction process without artificial selection. In 2006, Hinton and others [14] proposed using DBN to achieve nonlinear high dimensional data dimension reduction and classification, and triggered a wave of study of deep learning. Since then, deep learning as a new research direction of machine learning, has achieved great success in the [15-18] image, voice, text classification and pattern recognition etc. The study [19] proposed using DBN for aerial remote sensing image of road detection, and applied deep learning model to the field of spectral data analysis.

Depth network can fully exploit the relationship between hyperspectral image data, can get the deep relationship between the data, and can be used to establish a robust target feature extraction and classification recognition method. This paper attempts to apply the deep belief network to proceed the hyperspectral image data, combining with the inherent characteristics of hyperspectral image data, to explore hyperspectral image information processing method using deep learning method, with a view to hyperspectral image data nonlinear feature extraction is introduced a new method, and hope to introduce new methods for non-linear feature extraction for the hyperspectral image data.

II. DEEP BELIEF NETWORK MODEL

Hinton proposed deep learning model based on multilayer restricted Boltzmann machine with the energy function [20-22], the model is composed with multilayer structure from the bottom layer to the top layer [14]. Hidden layers of the DBN model are obtained through the weight and bias calculations, and the top-level units normalize output value to the probability using softmax activation function, finally the DBN gets the nonlinear mapping of the input data and the labeled data, and the algorithm regards the deep neural network as a binding of multilayer RBM [13]. The DBN model is a multi-layered learning model and can be trained by unsupervised and tuned by supervised and shown in Figure 1.

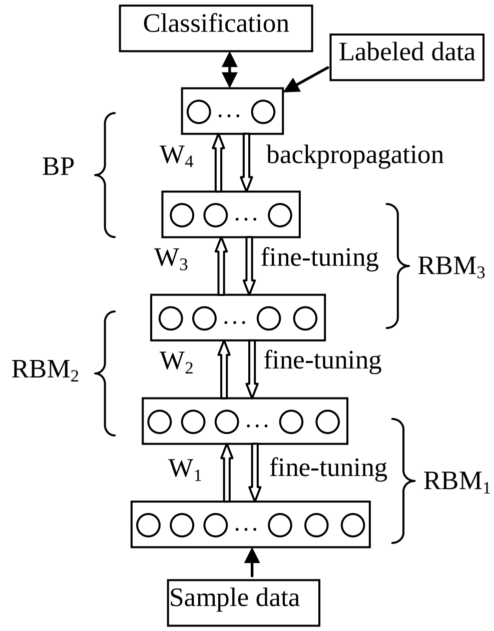

Figure 1. structure of DBN model

The DBN model uses the Wake-Sleep algorithm thinking [23], and the whole process is divided into pre-training and fine-tuning two stages. The DBN trains data for next layer in accordance with the bottom-up order in pre-training, and gets better performance weights through step by step learning. In fine-tuning, the DBN model regards the original training data as the supervised data, the pre-trained weight as the initial value, using gradient decreased algorithm and the back-propagation algorithm, to fine-tune the DBN according to the top-down sequence, and make DBN model achieve the optimal. Since the two stages of the whole process of DBN are linear in time and space complexity, therefore during training large quantities of data, DBN model is very effective.

a. Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM) is composed with a visible layer and a hidden layer, and the hidden layer is a feature detector, the RBM is shown in Figure 2. The nodes of the visible layer and the hidden layer of DBN are connected with each other, and the nodes the visible layer and the hidden layer are not connected with each other, so that the value of the same layer will not affect the probability distribution of each other.

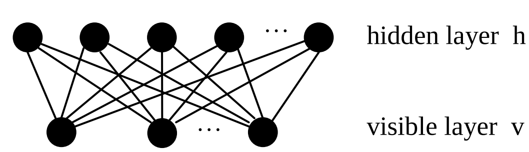

Figure 2. structure of RBM model

Assuming a RBM has n visible units and m hidden units, i represents the i-th visible unit, j represents the j-th hidden layer unit, vi represents the i-th visible unit state, hj represents the j-th hidden layer unit state, The energy function of the combined configuration of RBM is defined as the formula (1).

$$E(v,h|\theta) = -\sum_{i,j} v_i h_j w_{ij} - \sum_i v_i a_i - \sum_j h_j b_j \qquad (1)$$

Among them, $\theta = (a_i, b_j, w_{ij})$ is the RBM model parameter, w_{ij} is the connection weight between the visible layer unit i and the hidden layer unit j, a_i is the bias of the visible layer unit i, and b_j is the bias of the hidden layer unit j. Assuming the (v,h) satisfies the Boltzmann distribution, and bases on the energy function $E(v,h|\theta)$, the joint probability distribution of a configuration is obtained by $P(v,h)$ formula (2),

$$P(v,h|\theta) = \frac{e^{-E(v,h|\theta)}}{Z(\theta)} \qquad (2)$$

Among them, Z is the partition function, also known as the normalization factor, $Z(\theta) = \sum_{v,h} e^{-E(v,h|\theta)}$.

Because the RBM is a two part graph model, In the case of the visible layer unit v, the independent units in the hidden layer h satisfy the formula (3),

$$p(h|v,\theta) = p(h_1|v,\theta) = p(h_2|v,\theta) = \ldots = p(h_n|v,\theta) \qquad (3)$$

Similarly, the visible layer units satisfies the formula (4),

$$p(v|h,\theta) = p(v_1|h,\theta) = p(v_2|h,\theta) = \ldots = p(v_n|h,\theta) \qquad (4)$$

Therefore, when the RBM model input is the visible layer v, the hidden layer h is obtained by calculating the $p(h|v,\theta)$, and then the visible layer is reconstructed by $p(v|h,\theta)$. DBN model fines tune the parameters through the BP algorithm, so that the reconstruction of the visible similar to visible layer v, and the hidden layer unit h is the characteristics data of the input unit of visible layer.

Because there is no relationship between the pre-training process of RBM and dimensions of data, so the characteristics can be used to the effective projection of high-dimensional data, which is the reason why the DBN model is used to extract the feature bands of hyperspectral data.

b. Contrast Divergence Gradient Approximation

The activation probability formula (5) of hidden layer unit j is known when the input state of the visible unit is given by the structure of connection between layers and connectionless inter-layer of RBM.

$$p(h_j = 1 | v, \theta) = \sigma(b_j + \sum_i v_i w_{ij}) \qquad (5)$$

Among them, $\sigma = \frac{1}{1+\exp(-x)}$ is the sigmoid activation function. On the contrary, when the state of the hidden layer unit is given, the activation probability of the visible layer unit i is the formula (6),

$$p(v_i = 1 | h, \theta) = \sigma(a_i + \sum_j h_j w_{ij}) \qquad (6)$$

RBM training goal is to get $\theta = (a_i, b_j, w_{ij})$, and make the joint probability distribution $P(v,h)$ maximum [24], and achieve the purpose of fitting the given training data. Parameter θ can be got by maximizing the log likelihood function learning of the RBM in the training set, but the joint probability distribution $P(v,h|\theta)$ of visible layer unit v and hidden layer unit h is difficult to obtain, and can get approximation by Gibbs sampling [25]. In 2002, Hinton presented contrasting divergence (CD) algorithm to achieve RBM fast learning [26].The formula (5) is used to calculate the binary state of the hidden layer unit, and to determine the characteristic data retained in the hidden layer. The formula (6) is used to determine the probability of 1 of the visible layer, and then to obtain a reconstruction of the visible layer. The RBM parameters are updated by maximize the log likelihood function of stochastic gradient ascent method with reconstruction data of visible layer and the original visible layer data, the main basis of the weights updating is the correlation difference between the hidden layer activation unit and the visual layer inputs, as in equation (7-9).

$$w_{ij}^{new} = w_{ij}^{old} + \varepsilon(< v_i h_j >_{data} - < v_i h_j >_{recon}) \qquad (7)$$

$$a_i^{new} = a_i^{old} + \varepsilon(< v_i >_{data} - < v_i >_{recon}) \qquad (8)$$

$$b_j^{new} = b_j^{old} + \varepsilon(< h_j >_{data} - < h_j >_{recon}) \qquad (9)$$

Among them, ε is learning rate of pre-trained, $< \bullet >_{data}$ is the average value defined by the distribution of training data set, $< \bullet >_{recon}$ is a replacement of expect value of $< \bullet >_{model}$ calculated with contrastive divergence algorithm proposed by Hinton, and $<\bullet>_{recon}$ is result of 1-step Gibbs sampling or average obtained by repeating Gibbs sampling. The contrasting divergence algorithm makes it run only a small amount of Gibbs sampling to achieve weight updating, thereby speeding up the training process.

c. Fine-tuning of DBN model

Seen from the above RBM training process, the function of getting $<\bullet>_{recon}$ is Differentiable function using contrasting divergence algorithm in RBM, therefore, DBN model calculate error of gradient using back-propagation algorithm with the labeled data and regard squared error as the cost function, fine tune the model parameters, and achieve optimizing the DBN model role. The square error is calculated as the formula (10),

$$C = \frac{1}{2}\sum_{i}^{n}\sum_{j}^{m}(d_{ij} - y_{ij}) \qquad (10)$$

Among them, d is expected to output, y is the results got with contrasting divergence algorithm of the RBM, n is the number of data samples for training, m is the number of output units. According to the gradient descent algorithm, the RBM parameter change is calculated as the formula (11),

$$\Delta\theta = -\eta\frac{\partial C}{\partial\theta} \qquad (11)$$

Among them, η is the learning rate, the formula (11) is transformed to formula (12) using SoftMax regression model to calculate and obtain $\Delta\theta$,

$$\Delta\theta_{ij} = \eta\sum_{p}^{n}(d_{pi} - y_{pi})y_{pj} \qquad (12)$$

In comparison, the shallow layer neural network is easy to fall into local minimum because of the initial value of parameter setting. At the same time, in the DBN model, because more reasonable initial values are found in the pre-training process, it is better to use the back propagation algorithm to fine tune the parameters. BP algorithm in DBN model requires only a local search for the weight parameter space, training is speeded up, and the convergence time is also less.

III. EXPERIMENTAL DATA AND PARAMETER SETTING

a. experimental data and preparation

In order to verify the effectiveness of the DBN deep learning model implementing hyperspectral data dimensionality reduction, in this paper the AVIRIS hyperspectral original image experiments were carried out, and the image was taken from the northwestern Indiana India Pine remote sensing experiment area in June 1992 shooting. The original image has 220 bands, and the wavelength range is 400 ~ 2500nm, spectral resolution is 10nm, the image size is 145 * 145 pixels and a pixel depth is 16 bits, spatial resolution is 3.7m. The data set consists of 16 ground truths, the paper uses the 50, 27 and 17 bands spectral image to make the R, G, B false color image, and the false color image of Indian Pine is shown in Figure 3.The spectral data of 16 kinds of ground truth are used as the sample set, and the true value of the sample set is shown in Figure 4.

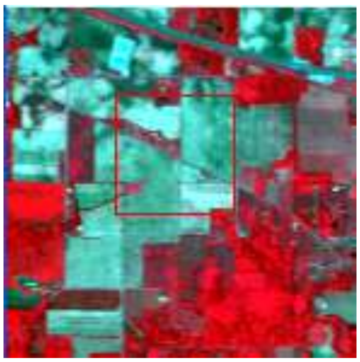

Figure 3. Original Indian Pine Hyperspectral Image

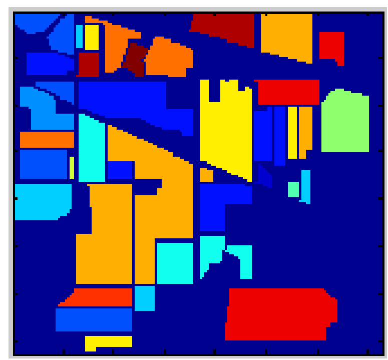

Figure 4. Indian Pine Ground Truth Graph

The experimental sample set is shown in table 1.

Table 1: Indian Pine Sample Set

#	Ground truth	Sample size	#	Ground truth	Sample size
1	Alfalfa	46	9	Oats	20
2	Corn-notill	1428	10	Soybean-notill	972
3	Corn-mintill	830	11	Soybean-mintill	2455
4	Corn	237	12	Soybean-clean	593
5	Grass-pasture	483	13	Wheat	205
6	Grass-trees	730	14	Woods	1265
7	Grass-pasture-mowed	28	15	Buildings-Grass-Trees-Drives	386
8	Hay-windrowed	478	16	Stone-Steel-Towers	93

The water vapor absorption bands, ozone effects bands, and low SNR bands are removed from image in this experiment, and noise ratio relatively high bands are retained to constitute the data set.

The 145*145*200 hyperspectral data set after pretreatment reformed 21025*200 two-dimensional data set through one-dimensional spectral data vectorization, the columns of the two-dimensional data set are 200 spectral bands, and seen as variables for feature extraction of hyperspectral image, and the rows values of the two-dimensional data set are hyperspectral data of the image.

b. experimental parameter setting

In this paper, we use matlab2015b language to achieve hyperspectral data feature extraction algorithm, the algorithm is debugged and run in the ThinkStation Lenovo image processing workstation, the operating system is Win 7. The DBN model parameters for achieving spectral data feature extraction are shown in table 2.

Table 2: DBN Model Parameter for Our Experiment

Parameters	Values
Number of batches	1
Number of nodes in input layer	200
Pre-training epoch	50
Learning rate for weights	0.001
Learning rate for biases of visible units	0.001
Learning rate for biases of hidden units	0.001
Weight of cost	0.0002
Initial momentum	0.5
Final momentum	0.9
Fine-tuning epoch	200

After removal of the low SNR band, the remaining 200 bands of the Indian pine data do for the experimental data, in order to achieve the purpose of extraction feature bands, the DBN model input layer nodes is 200, with the same number of spectral bands.

In this paper, the batch number of the input of the DBN model is set to 1, and the model parameters are updated by the way of online learning.

Due to the DBN model the training process, if the learning rate is too large, the DBN will lead to a rapid increase in reconstruction error, abnormal weight will become larger, therefore the initialization learning rate of the connection weights is set to 0.001.

If the learning rate is too large in model training process, the DBN model will not be stable, the stability of the model can be increased by setting a small learning rate, but the speed of convergence will be slow. In order to increase the stability of the DBN model, this paper proposes to use the momentum to realize updating the parameters, and achieves increasing the stability of the DBN model. The initial learning momentum is 0.5 in this paper experiments, and the momentum is 0.9 when the reconstruction error is stable. The updating of the momentum of the model is shown in Formula (13),

$$\theta_{ij}^{new} = \Delta p \theta_{ij}^{old} + \varepsilon(<\bullet>_{data} - <\bullet>_{recon}) \qquad (13)$$

The momentum learning rate Δp is increased based on the formula (7), (8), (9), and the range of the value of Δp is [0.5, 0.9].

RBM is the core of the model of DBN deep structure, and the accuracy of its training results increases with the increasing of the number of layers, but too much RBM layers will lead to overfitting to the whole DBN model in the training process. In order to avoid overfitting, the method is to increase the penalty in the process of updating the weight parameter w_{ij}. In this paper experiments, the weight of the penalty is updating as the formula (14),

$$w_{ij}^{new} = \Delta p w_{ij}^{old} + \varepsilon(<\bullet>_{data} - <\bullet>_{recon}) - \alpha w_{ij}^{old} \qquad (14)$$

Among them, α is the punishment coefficient, as shown in Table 1, the value is 0.0002.

During the experiment, the number of pre-training iteration of each RBM is 50 times, and the iteration number of the fine-tuning of DBN model is 200 times.

IV. ANALYSES OF THE RESULTS

The DBN model trains each layer RBM of it through unsupervised training methods firstly, and implements feature points of the sample data mapping to different feature spaces, as far as possible retaining the characteristic information of the original data. After that the supervised learning method used to train the whole network, and the DBN model parameters are optimized. In the process of calculation, the error caused by the unsupervised or supervised process will affect the final result of the model. The RBM is the core of the DBN model, different number of

RBM layers result in different model error. At present, there is no sound theoretical basis for selection of the RBM layer number of DBN model, and most of the researches about the depth of RBM selection were proposed by the prior knowledge [27, 28], it is difficult to give full play the advantages of deep DBN model. The paper [29] made it clear to the DBN deep model, an increasing the number of RBM layers can improve DBN model capabilities, but too more RBM layers will lead to a decline in DBN model overall generalization ability [30], DBN model structure should be specific research issues.

a. The influence of RBM layers

In this paper, the optimal network structure of the DBN model identified after much experimenting. To reduce the influence of non-related parameters on the training results, the influence of RBM layer on the results of DBN model is studied by assuming each hidden layer having the same number of nodes. In the experiment, the number of RBM layer of the DBN model is set to 1 layer to 7 layers corresponding to seven kinds of DBN structure model. The hidden layer units are 600 each structure, and the output layer units are 5, that is, after DBN dimension reduction, 5 characteristic bands are generated.

Hinton proposed the reconstruction error to evaluate the DBN model in the literature [25]. The reconstruction error is different between original data and the mapping results obtained by the RBM training input data, and the training data is used as the initial state (typically assessed by a two-norm or norm). Reconstruction error can reflect to some extent the likelihood of the training data and the original data, but it is not entirely reliable [25]. Therefore, in this paper, a number of experimental methods are used to analyze the influence of RBM layer on the DBN model, in which the RBM structures are set from 1 layer to 7 layers, and repeat the experiment 5 times for each structure. Figure 5 is characterized by obtaining different depth reconstruction error of reconstruction results of DBN model feature data and the original data.

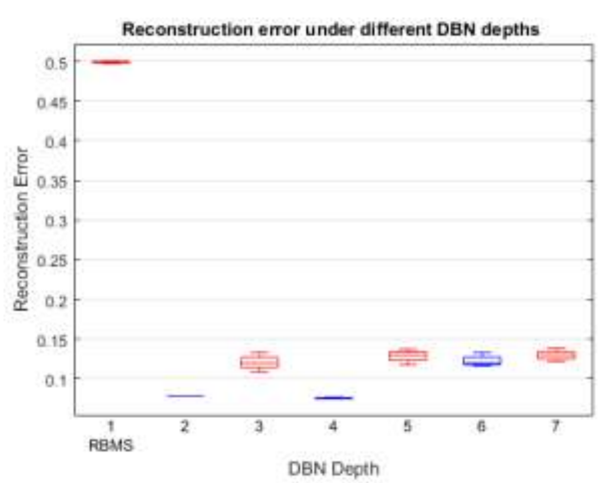

Figure 5. Reconstruction error with different depths of DBN

Figure 5 shows, when the RBM in DBN model had 1 layer, the reconstruction error was maximum, reached 0.5, the results of the 5 experiments did not differ much, The DBN model of 1 layer RBM is relatively stable, but the reconstruction errors were larger. The layer number of RBM in DBN model was 4, the reconstruction error was minimum, was 0.072, and the results gaps between the 5 experiment was small, which indicated that the reconstruction error of 4 RBMs in DBN model was minimum, and the model was most stable. The reconstruction error of the DBN model is gradually increasing from 5 RBM layers, and the results gaps between the 5 experiments were bigger, which indicated that the generalization ability of the DBN model beginning to decline from 5 layers, the over fitting phenomenon occurred, and DBN model began to become unstable. Therefore, the RBM layer number in the DBN model has great influence on the nonlinear low-dimensional projection of high-dimensional data, and the suitable RBM dimensions can improve the generalization ability of the DBN model.

Figure 6 is the reconstruction error of one of 5 experiments of bottom layer RBM of different depths DBN model.

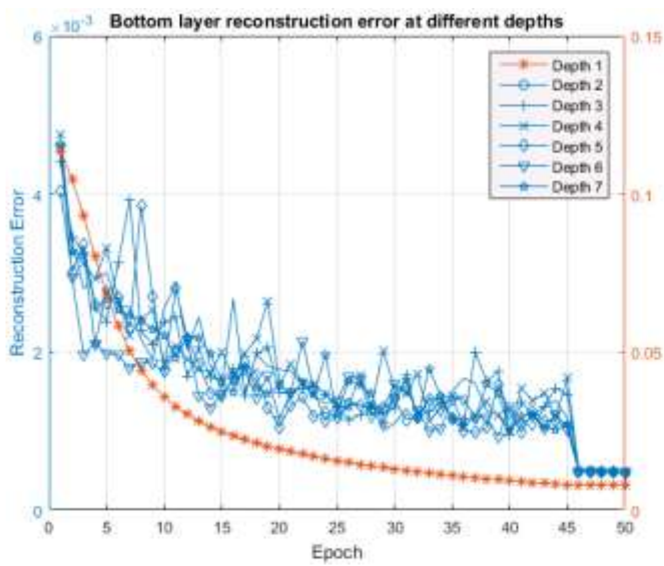

Figure 6. Reconstruction error with the bottom layer of different depths of DBN

From Figure 6 shows that with 1 RBM layer, and 50 times iterative training, the reconstruction error of underlying RBM decreased gradually from 0.11, finally reaches 0.008, and the reconstruction error changed relatively smooth in whole training process, which indicates that the training process of the RBM is relatively stable throughout. When the RBM layer number changed from 2 to 7, and iterative number was 50 times in training process, the reconstruction error of underlying RBM reduced from 0.005 to 0.0004, initial training error is smaller than 1 RBM layer and the number of iterations reached 45, the reconstruction error tended to be stable, but the reconstruction error change is not smooth in whole training process, which indicates that the underlying RBM was instable in the training process, did not reach the best state, and needs to increase the number of RBM.

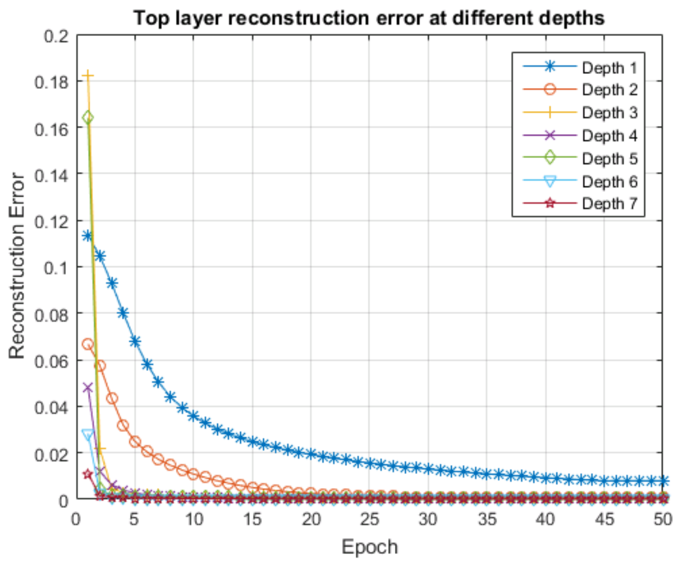

Figure 7. Reconstruction error with the top layer of different depths of DBN

Figure 7 shows the reconstruction error curve of top-level layers of different RBM layers in training process. Seen from the chart, the reconstruction error of top-level layer RBM decreased in the training process, and the reconstruction error curves were relatively smooth, which indicates that the model will tend to be stable in the top-level layer RBM in different depth DBN model. However, the reconstruction error is relatively large when the start of training for each top -level RBM, which was due to the initial weights are randomly set, so the reconstruction error is relatively large at beginning of trainings. Therefore, it is worth further research on how to initialize each layer of RBM, which is helpful to improve the overall performance of the model.

b. The influence of the hidden layer units

The paper [30] pointed out that different hidden layer unit number have a significant impact on the dimensionality reduction performance of the entire DBN network, when the hidden layer unit size is too small, connections between neurons correspond less, and not enough to extract the classification of relevant information, and the number of hidden layer units is too much also easily lead to over fitting problem. Hyperspectral data is highly redundant data, and the hidden layer can be set a little bit of units. In the experiment, the number of RBM layers was 4, the number of each hidden layer units was [600, 400, 200, 100, 50, 5], and the other parameters are fixed. For 600 hidden layer units, the deep structure of DBN is 200-600-600-600-5, 5 is the true dimension of the hyperspectral data through features after extraction. As can be seen from figure 8, when the number of unit in the hidden layer is 100, the reconstruction error of the DBN model is the smallest, but the operation cost is a little bit longer.

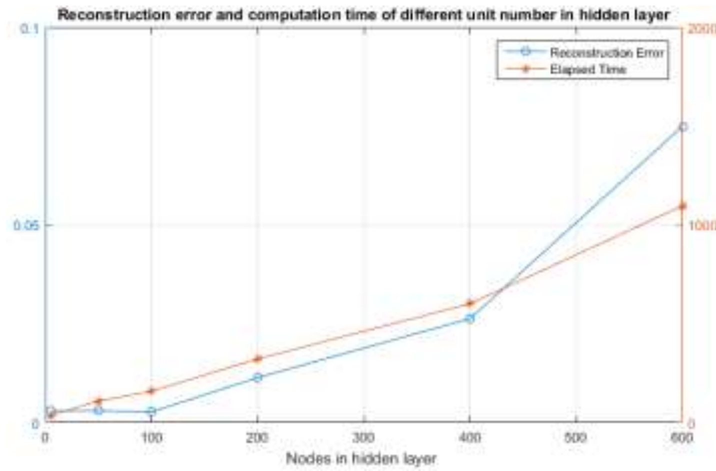

Figure 8. Reconstruction error and computation time of different number of nodes in hidden layer

In order to verify the validity the effectiveness of using DBN model to extract the feature of hyperspectral data. In this paper, the result of the feature extraction with different hidden layer unit size of DBN model were used as the input of SVM classifier[31,32], and the indina pine hyperspectral data was used to classify. The SVM algorithm used the open source tool LibSVM as multi-classifier, and used the Gauss RBF kernel function as SVM kernel function. Penalty coefficient C and the intervals are two parameters RBF kernel function must be set, use the grid search method of cross-validation to determine in experiments.

Figure 9 is the classification results obtained by setting the penalty coefficient of Gauss RBF kernel function C=50 and γ =0.2 with different units in the hidden layer in several experiments.

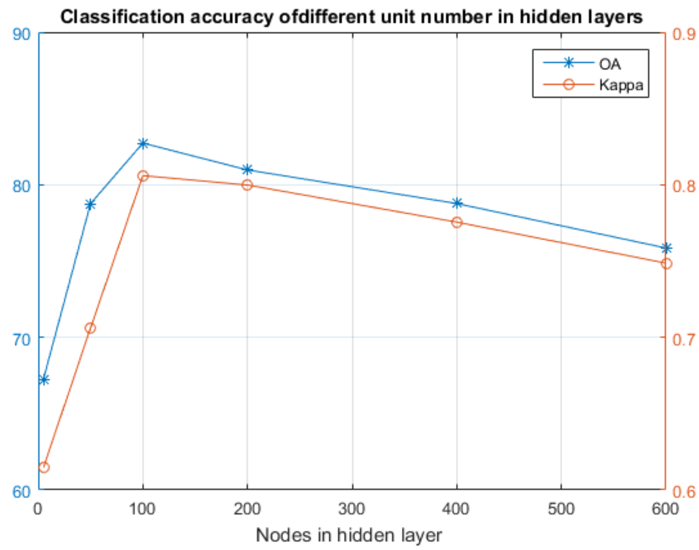

Figure 9. Classification accuracy of different number of nodes in hidden layer

It is easy to see that the DBN model in the deep structure of 200-100-100-100-5, the classification of the overall accuracy (OA) and the Kappa coefficient is the largest. The overall accuracy (OA) and the Kappa coefficient of is the second when the number of hidden layer units is 200. Table 3 for the hidden layer node number is 100, the SVM classification results of the confusion matrix.

Table 3 shows the classification confusion matrix using SVM as classifier when hidden layer units are 100.

Table 3: Confusion matrix of classification results for 100 hidden layer nodes

Class	**Ground-truth**															
	#1	**#2**	**#3**	**#4**	**#5**	**#6**	**#7**	**#8**	**#9**	**#10**	**#11**	**#12**	**#13**	**#14**	**#15**	**#16**
#1	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#2	0	1057	0	6	0	0	0	0	0	13	0	0	0	0	0	5
#3	0	21	731	0	0	0	0	0	0	0	0	0	0	0	0	0
#4	0	84	9	214	8	0	0	0	0	44	37	14	0	41	0	0
#5	0	53	0	0	435	36	0	2	0	0	0	0	0	45	2	0
#6	1	0	0	2	0	567	0	4	2	0	36	1	4	68	0	0
#7	0	0	0	0	3	0	26	7	0	12	68	0	0	0	0	0
#8	3	52	0	0	7	32	2	413	0	2	23	0	0	7	11	0
#9	0	0	0	0	27	13	0	25	15	0	80	21	13	0	16	0
#10	0	68	37	0	0	0	0	24	0	851	0	41	0	13	0	2
#11	0	20	30	0	1	0	0	3	0	0	1963	0	0	57	12	2

#12	0	52	23	12	2	47	0	0	0	0	123	516	0	0	0	0
#13	0	21	0	2	0	0	0	0	3	32	74	0	186	0	0	0
#14	0	0	0	0	0	35	0	0	0	0	51	0	0	1034	0	0
#15	0	0	0	1	0	0	0	0	0	18	0	0	2	0	345	0
#16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	84
Total	46	1428	830	237	483	730	28	478	20	972	2455	593	205	1265	386	93
Accuracy	0.91	0.74	0.88	0.90	0.90	0.78	0.93	0.86	0.75	0.88	0.80	0.87	0.90	0.82	0.89	0.90

In this paper, the DBN model is used to extract the feature bands of hyperspectral data, and used the SVM algorithm to classify in low dimension feature space, and got good classification result finally, which stated that the inherent law of hyperspectral data were better explored by multi-layer RBM in DBN through unsupervised learning and supervised fine-tuning, made a good foundation for the latter to effectively classify and identify.

V. CONCLUSIONS

Hyperspectral data dimensionality reduction is the first step in hyperspectral image analysis and application, and is an important means to obtain spectral information. In this paper, a new method based on depth learning model, DBN model, is proposed to extract the hyperspectral data. Compared with the traditional BP neural network, the DBN is adopted based on statistical methods, through combination low-level features of the data sample, forms more high-level abstraction represents of feature, and can discover the real relationship of the bands, and can get more accurate description of spectral space, and avoid the uncertainty of traditional neural networks applied to hyperspectral data.

In this model, each RBM of the DBN model is trained by unsupervised method, and the parameters of DBN model are obtained, then the BP network is used to optimize the parameters, and makes the DBN model to achieve the best, and finally the SVM classifier is used to achieve data classification. Through the experiment finds that the DBN deep structure model has strong scalability, if RBM layers number of the model is too little, the DBN model expression ability is insufficient, but over more RBM layers will make the model produce the phenomenon of over fitting, and proper DBN model structure can help to improve the accuracy of classification and recognition of hyperspectral data.

VI. ACKNOWLEDGMENT

This work was supported by national and international scientific and technological cooperation special projects (No. 2015DFA00530), and supported by national natural science foundation of China (No.61461042, No.61461041).

REFERENCES

[1] JIA X, KUO B C, and CRAWFORD M M. Feature mining for Hyper-spectral image classification[J]. Proceedings of the IEEE, 2013, 101(3):676-697.

[2] Wang Liping. FEATURE SELECTION ALGORITHM BASED ON CONDITIONAL DYNAMIC MUTUAL INFORMATION [J].INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS, 2015, 8(1):316-337

[3] HUO LEI-GANG, FENG XIANG-CHU. Denoising of Hyperspectral Remote Sensing Image Based on Principal Component Analysis and Dictionary Learning[J]. Journal of Electronics & Information Technology,2014,36(11):2723-2729.

[4] CHENG SHU-XI, XIE CHUAN-QI, WANG QIAO-NAN, et al. Different Wavelengths Selection Methods for Identification of Early Blight on Tomato Leaves By Using Hyperspectral Imaging Technique[J]. Spectroscopy and Spectral Analysis,2014,34(5):1362-1366.

[5] FAN LIHENG, LV JUNWEI, and DENG JIANGSHENG. Classification of Hyperspectral Remote Sensing Images Based on Bands Grouping and Classification Ensembles[J]. ACTA OPTICA SINICA,2014,VOl.34,No.9:1-11.

[6] CHERIYADAT A, BRUCE L. Why principal component analysis is not an appropriate feature extraction method for hyperspectral data[J].Proceeding of IEEE Geoscience and Remote Sensing Symposium(IGARSS),2003,104(2):3420-3422.

[7] SUN KANG, GENG XIURUI, TANG HAIRONG, et al. A New Target Detection Method Using Nonlinear PCA for Hyperspectral Imagery [J].Bulletin of Surveying and Mapping, 2015(1):105-108.

[8] LIU JING. Kernel Direct LDA Subspace Hyperspectral Image Terrain Classification[J].Computer Science,2012,39(6): 274-277.

[9] DC FENG, F CHEN, and XU WEN-LI. Detecting Local Manifold Structure for Unsupervised Feature Selection [J]. Acta Automatica Sinica, 2014, 40(10):2253-2261.

[10] S ZHOU, K TAN, and L WU. Hyperspectral Image Classification based on ISOMAP Algorithm using Neighborhood Distance[J].Remote Sensing Technology & Application, 2014, 29(4):695-700.

[11] L YAN, DP ROY. Improved time series land cover classification by missing-observation-adaptive nonlinear dimensionality reduction[J].Remote Sensing of Environment, 2015, 158:478-491.

[12] W SUN, C LIU, B SHI, et al. Dimensionality Reduction with Improved Local Tangent Space Alignment for Hyperspectral Imagery Classification [J].Journal of Tongji University, 2014, 42(1):0124-0130.

[13] HINTON G, OSINDERO S, and TEH Y W. A fast learning algorithm for deep belief nets [J].Neural computation, 2006, 18(7):1527-1554.

[14] HINTON, GEOFFREY E, and RUSLAN R. Reducing the Dimensionality of Data with Neural Networks [J].Science, 2006, 313(5786):504-507.

[15] YANGYAN LI, HAO SU, CHARLES RUIZHONGTAI QI, et al. Joint embeddings of shapes and images via CNN image purification [J]. Acm Transactions on Graphics, 2015, 34(6):1-12.

[16]HINTON G, DENG L, YU D, et al. Deep Neural Networks for Acoustic Modeling in Speech Recognition [J]. IEEE Signal Processing Magazine, 2012, 29(6): 82-97.

[17]DAHL G E, YU D, DENG L, et al. Context-Dependent Pre-trained Deep Neural Networks for Large-Vocabulary Speech Recognition [J]. IEEE Trans on Audio, Speech and Language Processing, 2012,20(1):30-42.

[18] J XIE, L ZHANG, J YOU, et al. Effective texture classification by texton encoding induced statistical features [J]. Pattern Recognition, 2015, 48(2):447–457.

[19]MNIH V, HINTON G E. Learning to Detect Roads in High-Resolution Aerial Images [J]. Lecture Notes in Computer Science, 2010, 6316:210-223.

[20] HINTON G E，SEJNOWSKI T J. Optimal perceptual inference[C].Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Piscataway, NJ:IEEE Press, 1983: 448-453.

[21] RD HJELM, VD CALHOUN, R SALAKHUTDINOV, et al. Restricted Boltzmann machines for neuroimaging: an application in identifying intrinsic networks[J]. Neuroimage, 2014, 96(8):245-260.

[22] LE ROUX, NICOLAS, YOSHUA BENGIO. Representational power of restricted boltzmann machines and deep belief networks[J]. Neural Computation, 2008, 20(6): 1631-1649.

[23] J BA, R GROSSE, R SALAKHUTDINOV, et al. Learning Wake-Sleep Recurrent Attention Models[J]. Conference on Neural Information Processing Systems, Canada, 2015:1-9.

[24] YOSHUA BENGIO, PASCAL LAMBLIN, DAN POPOVICI, et al. Greedy layer-wise training of deep networks[C]. Advances in Neural Information Processing Systems 19 (NIPS 2006).Vancouver, 2007: 153-160.

[25] HINTON G E. A Practical Guide to Training Restricted Boltzmann Machines[R], Montreal: Department of Computer Science, University of Toronto, 2010

[26] HINTON G E. Training products of experts by minimizing contrastive divergence[J]. Neural Computation, 2002, 14(8):1771-1800.

[27] LE ROUX, NICOLAS, YOSHUA BENGIO. Representational power of restricted boltzmann machines and deep belief networks [J]. Neural Computation, 2008, 20(6): 1631-1649.

[28] ZHANG CHUN-XIA, JI NAN-NAN, and WANG GUAN-WEI. Restricted Boltzmann Machines[J]. Chinese Journal of Engineering Mathematics, 2015, 32(2):159-173.

[29] RN LE, Y BENGIO. Representational power of restricted boltzmann machines and deep belief networks [J]. Neural Computation, 2008, 20(6):1631-1649.

[30] H LAROCHELLE, Y BENGIO, LOURADOUR, et al. Exploring Strategies for Training Deep Neural Networks [J]. Journal of Machine Learning Research, 2009, 10(6):1-40.

[31] Yongqing Wang,Yanzhou Zhang. NOVEL MULTI-CLASS SVM ALGORITHM FOR MULTIPLE OBJECT RECOGNITION[J].INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS,2015,8(2):1203-1224

[32] Yongqing Wang, Xiling Liu. FACE RECOGNITION BASED ON IMPROVED SUPPORT VECTOR CLUSTERING[J].INTERNATIONAL JOURNAL ON SMART SENSING AND INTELLIGENT SYSTEMS,2014,7(4):1807-1829