DBM Optimization with Additional Category Information

Kai Liu, Jie Zhang, Xinghai Wang
Department of Aviation Foundation, Naval Aeronautical University, Yantai 264001, China
wendao_2008@163.com

Abstract. For unconstrained training of restricted Boltzmann machine (RBM), it is easy to appear that feature homogenization leads to poor generalization ability. This paper introduces category conditions into DBM training, proposes label conditional RBM which is used in the construction of DBM. According to unsupervised training characteristics of RBM, this paper adds the category information as the model hidden unit training condition to the implicit unit posterior activation probability calculation. This paper applies the model as the underlying structure of the deep Boltzmann machine (DBM) to the deep network construction. Through handwritten digit recognition set test, compared with the shallow model, the new model after adding the category condition has a great improvement in the model training speed and feature extraction effectiveness, and can effectively enhance the feature learning ability of the deep model.

1. Introduction
As a structure of the artificial neural network, Restricted Boltzmann Machine (RBM) based on the energy model has the advantages of simple network form and fast model convergence. Therefore, RBM has been widely used in machine learning problems such as text clustering[1,2], image denoising[3], handwritten digit recognition[4], speech recognition[5], and collaborative filtering[6]. A new field has been formed - deep learning[7].

However, the training of RBM belongs to unsupervised learning. The existence of model training relies solely on unlabelled data, which are prone to feature homogeneity, which leads to poor generalization ability of the model. Therefore, the RBM model structure is improved, and the tag information is introduced into the RBM training. Designing an RBM combined with tag information can further accelerate the model training and improve the generalization ability of the model data.

The category information processing on the basis of RSM was added[8], and a supervisory-based RSM-sRSM was proposed, but there was a problem of category information reconstruction, which leads to training divergence. Therefore, based on RBM, a class condition-based RBM (label condition RBM, ICRBM) is designed. For the sRSM defect, the category information is used as the RBM training parameter update condition, which avoids the problem that the category information needs to be reconstructed. The experimental results show that the features of RBM extraction based on category conditions are better for classification and the training efficiency is also improved.

2. Restricted Boltzmann Machine
Restricted Boltzmann Machine is based on the Boltzmann Machine (BM). By limiting the intra-layer unit connection in the BM, the activation probability conditions of adjacent layer units are independent in a given unit state. As an undirected graph model, the visible cell layer in RBM represents
observation data; the hidden cell layer is represented as a feature detector, and its structure is shown in Fig.1.

![Fig.1 RBM units connection diagram](image)

RBM training is achieved by maximizing the data likelihood probability and is unsupervised. In view of the difficulty in calculating the average expectation of model data during RBM training, the Contrast Divergence (CD) algorithm was proposed\[9\], using Gibbs sample values as the model expectation. The CD algorithm first performs the state transition by performing the Gibbs Sampling with each training data as the initial state; then, the transferred data is used as the sample to estimate the mean of the Negative Phase during the RBM training to implement the parameter update.

Experiments show that only one state iteration in the application can ensure the good learning effect of the model. Then, given the training data $v^{(n)}$, the connection weight $W_{ij}^{(n)}$ is updated as shown in equation (1).

$$
\Delta W_{ij}^{(n)} = P(h_j = 1 | v^{(n)})g_{h_j}^{(n)} - P(h_j = 1 | v^{(n-1)})g_{h_j}^{(n-1)}
$$

3. Restricted Boltzmann machine with label conditions

3.1. Model structure

Different from the category information\[8\] as an additional visible unit participation model, the lCRBM designed in this paper uses the label unit as the conditioning unit of the RBM. Therefore, the structure of the lCRBM model is shown in Fig.2.

![Fig.2 lCRBM units connection diagram](image)

It can be seen from Fig.2 that the lCRBM is composed of three units, which are visible unit $V$, hidden unit $H$ and category condition unit $L$; while the hidden unit layer, the visible unit layer, and the category condition layer are interconnected, and the same layer unit is not connected; $W$ represents the connection weight between the hidden unit layer and the visible unit layer, and $D$ represents the connection weight between the hidden unit layer and the category condition layer.

Then, given the state of each unit $(v, h, 1)$, the energy definition of the lCRBM model is as shown in equation (2).
\[ E(v, h, l) = -\sum_{i=1}^{N} v_{i} h_{l} - \sum_{j=1}^{M} h_{j} c_{j} - \sum_{i=1}^{N} \sum_{j=1}^{M} W_{ij} v_{i} h_{j} - \sum_{i=1}^{L} \sum_{j=1}^{M} D_{ij} l_{i} h_{j} \]  \tag{2}\]

\( l \) is the category condition unit, \( D_{ij} \) is the connection weight between the hidden unit \( h_{j} \) and the category condition unit \( l_{i} \), and \( L \) is the total number of category condition units.

Through the energy function of the ICRBM model, the posterior activation probability of the visible unit in the ICRBM can be derived, as shown in equation (3).

\[ p(v_{i} | h) = \sigma \left( \sum_{j=1}^{M} h_{j} W_{ij} + b_{i} \right) \]  \tag{3}\]

Similarly, the posterior activation probability of the hidden unit in ICRBM is as shown in equation (4).

\[ p(h_{j} | v, l) = \sigma \left( \sum_{i=1}^{N} v_{i} W_{ij} + c_{j} + \sum_{i=1}^{L} l_{i} D_{ij} \right) \]  \tag{4}\]

It can be seen that the hidden unit in ICRBM differs from the original unit is that class unit participates in the calculation of the posterior activation probability of the hidden unit, which is also ICRBM introduces the category information into the model training, thereby enabling training has a category-specific approach, weakening the feature homogenization problem that is easy to occur in RBM unsupervised training, and thus improving the data fitting degree.

### 3.2. Training method

The training goal of ICRBM is also to maximize the data likelihood probability, so the model training can be realized by CD algorithm. Since the category unit participates in the implicit unit posterior activation probability calculation, the training method of ICRBM is as follows:

For training data \( x=(x_{1}, x_{2}, x_{3}, ..., x_{n}) \) and corresponding category data \( l=(0,0,...,1,...,0) \), the ‘1’ value represents the category corresponding to the data.

Setting: learning rate \( \eta \), connection weight between hidden units and visible units \( W \), connection weight between hidden unit and category unit \( D \), visible unit offset \( b \), hidden unit offset \( c \)

(1) For all hidden units \( i \), calculating the posterior activation probability of the hidden unit based on equation (4).

(2) Sampling the state of the hidden unit according to \( Q(h_{i} = 1 | v, l) \).

(3) For all visible units \( j \), calculated \( P(x_{i,j} = 1 | h_{i}) \) according to equation (3).

(4) Sampling the state of the visible unit \( x_{i,j} \in \{0,1\} \) according to the posterior probability of the visible unit \( P(x_{i,j} = 1 | h_{i}) \).

(5) Calculating \( Q(h_{i} = 1 | X_{-}) \) by \( X_{-} \) and \( l \).

(6) Updating weight:

\[ W = W + \eta \left( H_{-} X_{-} - Q(H_{-} = 1 | X_{-}) X_{-} \right) \]

\[ b = b + \eta \left( X_{-} - X_{-} \right) \]

\[ c = c + \eta \left( H_{-} - H_{-} \right) \]

\[ D = D + \eta \left( I H_{-} - I Q(H_{-} = 1 | X_{-}) \right) \]

It can be seen from the training method of ICRBM that the category unit participates in the model training. As a feature extracted by the model, the hidden unit learns the combined characteristics of the initial data and the category data from the sample data and the corresponding category information during the training process.
3.3. Deep model implementation

The construction of the deep model usually takes the form of superimposing multiple RBMs and fine tuning the entire model parameters by BP algorithm. Therefore, the lCRBM model can be used as the underlying RBM, and the new deep model can be formed by superimposing the RBM.

A deep model lCDBM is constructed similarly to the deep Boltzmann Machine. We take the 3-layer structure as an example, the lCDBM model is shown in Fig.3.

![Fig.3 lCDBM units connection diagram](image)

4. Experimental results and analysis

In order to verify the performance of the proposed lCRBM model, the MNIST handwritten digital recognition set is selected as the experimental object[10]. The MNIST handwritten digit recognition set contains 10 handwritten digital images from 0 to 9 with a total of 70,000 images, all of which are 28×28.

Experimental setup: In order to analyze the feature extraction ability of lCRBM, two experiments are designed to test the accuracy of the features extracted by lCRBM as a shallow model for handwritten digit classification and the deep feature extraction ability as a deep structure DBM bottom model.

(1) Experiment 1

Experiment 1 is used to compare the effectiveness of RBM and lCRBM extraction features. In the experimental setting, the number of visible cells in the RBM is 784, the number of hidden cells is 200, the learning rate \( \eta = 0.01 \), and the number of cycles does not exceed 500. During the training process, the reconstruction errors of the two models are shown in Fig.4.

It can be seen from Fig. 4 that the reconstruction error of lCRBM is faster than RBM, indicating that lCRBM has a faster learning rate for training samples.

After completing the handwriting feature extraction, the radial basis support vector machine (RBF-SVM) and linear support vector machine (lSVM) provided by LIBSVM [10] are selected as the final classifier, and the parameters are all set by default. The implicit unit values of the RBM and lCRBM models are used as the input data of the RBF-SVM and lSVM, and the classification accuracy obtained are shown in Fig.5.
By adding category information, RBM can obtain more effective features, which improve the classification accuracy of the classifier. Both RBF-SVM and LSVM are consistent with this trend, and with the increase of the number of hidden cells, the performance of using LSVM classifier is better than using RBF-SVM classifier. This shows that the more hidden cells which the characteristics of RBM extraction, the more consistent with the linear distribution.

2) Experiment 2

Experiment 2 is used to test the influence of the basic structure of lCRBM as the deep model on the extracted features. Therefore, a 3-layer DBM is set as the experimental structure, in which the number of visible cells - the first hidden cells - the second hidden cells are 784-500-300, respectively. The experimental models are lCRBM-DBM (based on lCRBM) and standard DBM (based on standard RBM). The training method of the model is as described in[11], and RBF-SVM is also used as a classifier for handwritten numbers. The classification accuracy is shown in table 1.

| Model          | Classification accuracy |
|----------------|-------------------------|
| DBM            | 98.20%                  |
| lCRBM-DBM      | 99.00%                  |

It can be seen from table 1 that the accuracy of the DBM classification based on the lCRBM deep model is 99%, which are greatly improved compared with the standard DBM. Therefore, using lCRBM as the basic structure of the deep model can improve the feature extraction ability of the deep model and promoting the model which can greatly help the generalization of data.
5. Conclusion

This paper constructs an improved model of RBM-lCRBM, which implements the introduction of category information into DBM training. Compared with the standard RBM, the model can achieve model convergence and learning of training data faster, and the extracted features are more effective for classification due to the influence of category information. At the same time, the research is also based on ICRBM to test the basic depth model performance. The experimental results show that ICRBM is effective for handwritten digit recognition, and the idea of using category information as RBM training conditions can also be applied to fields such as text clustering and image recognition. The improvement of ICRBM in the future will also have a good guiding role in other fields of application.

Acknowledgments

This work was financially supported by the Taishan Scholar Special Foundation of Shandong Province (No.ts201511020).

References

[1] Mocanu D C, Mocanu E, Nguyen P H, et al. A topological insight into restricted Boltzmann machines[J]. Machine Learning, 2016, 104(2-3):243-270.
[2] Salakhutdinov R, Hinton G. An Efficient Learning Procedure for Deep Boltzmann Machines[J]. Neural Computation, 2012, 24(8):1967-2006.
[3] Pires R G, Santos D S, Souza G B, et al. A Deep Boltzmann Machine-Based Approach for Robust Image Denoising[C]// Iberoamerican Congress on Pattern Recognition. Springer, Cham, 2017.
[4] Papa, João Paulo UNESP, Scheirer W, Cox D D. Model selection for discriminative restricted boltzmann machines through meta-heuristic techniques[J]. Journal of Computational Science, 2015, 9:14-18.
[5] Ding W, Xiao L, Jing W, et al. Population activity changes during a trial-to-trial adaptation of bullfrog retinal ganglion cells[J]. Neureport, 2014, 25(10):801-805.
[6] Salakhutdinov R, Mnih A, Hinton G. Restricted Boltzmann machines for collaborative filtering[C]//Proceedings of the 24th international conference on Machine learning. ACM, 2007:791-798.
[7] Zrira N, Khan H A, Bouyakhf E H. Discriminative Deep Belief Network for Indoor Environment Classification Using Global Visual Features[J]. Cognitive Computation, 2018(1):1-17.
[8] Liu Kai, Zhang Limin, Zhang Jianing, etc. Improvement of RSM based on supervised learning[J]. Computer Engineering and Design, 2014, 35(4): 1429-1432(in Chinese).
[9] Zhang Jian, Ding Shifei, Zhang Nan, et al.. Incremental extreme learning machine based on deep feature embedded[J]. International Journal of Machine Learning and Cybernetics, 2016, 7(1):111-120.
[10] Chang C C, Lin C J. LIBSVM: a library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): 27-33.
[11] Wu J, Mazur T R, Su R, et al. A Deep Boltzmann Machine-Driven Level Set Method for Heart Motion Tracking Using Cine MRI Images[J]. Medical Image Analysis, 2018, 47:68-80.