American Journal of Engineering Research (AJER) 2017

American Journal of Engineering Research (AJER)

e-ISSN: 2320-0847 p-ISSN : 2320-0936

Volume-6, Issue-5, pp-186-194

www.ajer.org

Research Paper Open Access

Deep Multiconnected Boltzmann Machine for Classification

G. S.Tsanev

Department of Computer Systems and Technology Technical University of Gabrovo, Gabrovo, Bulgaria

ABSTRACT: *People want to operate mechanisms remotely using their own voice. In the epoch of humanoid robots the remote control of high technology products needs reliable classification tools. Classification algorithms are widely used for recognition of speech. There are a lot of machines with embedded voice control modules such as robots, wheelchairs, home systems etc. Comparing various kind of classifiers we assert that Boltzmann machine is a good choice. On the other hand the advantages of deep architectures are undoubted. A new multiconnected deep Boltzmann machine has been created and investigated in the present paper. The adaptive variable learning rate has been introduced for training the Boltzmann machine. An original LR search procedure is used for efficient error reduction. The new training algorithm is described in pseudocode. Comparisons between classification with and without LR search procedure is discussed.*

Nomenclature

$\bar{v}$	2-valent cepstral coefficient tensor of all frames representing a fixed word
$\underline{v}$	cepstral coefficient vector of a fixed frame
v	cepstral coefficient
$\bar{h}^{(j)}$	2-valent binary tensor of hidden units
$\underline{h}_i^{(j)}$	hidden unit (binary vector)
$\underline{u}$	vector of binary labels
P	Distribution
p	Likelihood
$\mathbb{H}$	3-valent binary tensor containing L hidden layers
m	number of visible units (frames)
L	number of hidden layers
s_k	number of units in the k-th hidden layer
n	number of classes of the output layer
t	number of training pairs
σ_i	standard deviation of the input unit $\underline{v}_i$
$w_{ij}, \widetilde{w}_{ij}$	weight matrices
$W, \widetilde{W}$	4-valent real-valued weight tensors or corresponding rectangular block weight matrices
$\underline{U}_{ij}^{(k)}$	weight vectors
$(\bar{v}, \underline{u})$	a training pair
$\mathbf{V}$	set of training pairs
$\bar{a}$	offsets
$\bar{b}, \bar{c}^{(j)}, \underline{d}$	biases
Θ	set of all biases, offsets and weights of the neural network
sigm	logistic function
Sigm	multivariate sigmoid
$\mathcal{N}$	multivariate Gaussian
λ	variable adaptive learning rate
ν	the smallest acceptable value for the module of the learning rate

I. INTRODUCTION

The classification of information is a very important activity in modern life. The contemporary data streams cannot be processed without using computer technology. The role of the neural networks is out of the question as it concerns classification of information from different areas in real life. The tendency to innovate in recent years has been the development of the humanoid robots which perform human-robot interaction. They implement facial expressions, some gestures, small talk, emotions [1], etc. Moreover, the robots can be used in order to ease daily life of ordinary people. A user can control a robot by means of a database with small number of commands. Each voice command corresponds to a sequence of actions which the robot should perform.

There are a lot of applications of Boltzmann machines used for classification in various areas. Emotion recognition from thermal images has been realized by S. Wang et al. [2]. Estimation of music similarity has been done by J. Schluter and C. Osendorfen [3]. Deep learning applications can be found in the area of biomedicine by Mamoshina [4] et al. G. Montavon [5] et al. consider deep Boltzmann machine (DBM) as a very efficient approach for extracting the structure of explored data. The authors introduce the term feed-forward in order to represent DBM which is trained with a real dataset and observing the evolution of theneural network layer after layer. They conclude that the idea of feed-forward hierarchy is better than the layer-wise approach.P. Xie [6] et al. have successfully applied diversifying RBM for document modeling. Semantic analysis is a very important tool for improving the classification success [7, 8, 9, 10, 11]. Latent semantic analysis [12, 13, 14, 15, 16, 17, 18] is used for data filtering and assigning to a certain class. J. Tomczak has outlined classification RBM for medical purposes [19]. The author shows that discriminative and sparse learning can be used for obtaining better results. P. Donovan has applied RBM as a powerful tool for recognizing some human gestures [20]. The author compares three methods for classification: Restricted Boltzmann Machines, k −nearest neighbors with principal component analysis and classical discriminative neural networks. He concludes that the usage of RBM significantly reduces the recognition error and generate much better results than the other classifiers. The depth of the network is very important feature. Deep architectures are composed of many hidden layers. They efficiently represent the input at different levels of abstraction. The algorithms based on deep architectures have advantages over the shallow one because they use less memory and less computer resources [21]. The Contrastive Divergence algorithm fails for the multilayer deep belief network [21] but in combination with mean field approximation it can be successfully used for training multilayer Boltzmann machines. The DBM is the most power tool for classification in speech recognition systems. It can be trained for a specific task using small amount of label data. The training of DBM with unlabeled data is layer-by-layer process [22, 23]. Voice control modules are embedded in a lot of applications that everybody uses in his daily life such as navigating systems, voice banking, drones, home systems, wheelchairs etc. The generative models on the basis of DBM present data structures more realistic [24]. An efficient learning of DBM with a layer-by-layer “pre-training” phase was published in 2010 by Salakhutdinov and Hinton [22].The idea is to initialize the weights sensibly. Many authors have applied DBM in practice for image retrieval [25], speech recognition [26], auto-encoders [27] etc.

A deep belief network has multiple hidden layers. The connections between the top hidden layers are undirected. In the lower layers the connections are directed. Their learning algorithm is greedy layerwise using the posterior distribution of the previous layer [22]. As opposed to DBNs, DBM are with undirected connections between all layers of the network. The training of the layers is more efficient because all of them are trained jointly not one by one [22]. In some DBN only the first two layers form an RBM which means in this case that an undirected model is realized. The other layers form a directed generative model. In DBM each pair of layers forms an RBM [23].

The last decades have witnessed the progress of automatic speech recognition applications. A stochastic model of voice control system has been obtained by Ilarionov [28] et al. The idea was extended by Todorov and Tsanev in [29] who consider DBM for controlling moving objects. The present paper is devoted to an L – hidden layer multiconnected classification Boltzmann machine. The DBM consists of one layer of visible units, L −layers of hidden units and an output layer. Each hidden layer is connected to all of the other hidden layers as well as the output layer. There are no horizontal connections between the hidden units in the same layer. An adaptive learning rate is used for the training process. On this basis a search procedure for error reduction is realized. A new architecture of a neural network has been created. It represents better the structure of the neurons in the human brain. The model has been implemented with MATLAB software product. A comparison of the training algorithm with variable and constant learning rate is presented.

Further, the paper is organized as follows. A new multiconnected DBM is described in Section 2. Algorithmic aspects of training the $L-$ hidden layer multiconnected Boltzmann machine are analyzed in the same section. The effect of learning rate (LR) search procedure is illustrated in Section 3. Concluding remarks are presented in Section 4.

II. DEEP MULTICONNECTED BOLTZMANN MACHINE WITH L-HIDDEN LAYERS

The proposed model of Boltzmann machine, Figure 1 is with a single visible layer, L hidden layers and a single output layer. The connections between the visible layer $\overline{v}$ and the first hidden layer $\underline{h}_j^{(1)}$ are given by weights w_{ij}. All hidden layers are connected each other. Additionally, they are connected to the output layer. The energy of the model is defined as follows:

$$E(\overline{v},\mathbb{H},\underline{u}) = \frac{1}{2}\sum_{i=1}^{m}\frac{(\underline{v}_i-\underline{a}_i)^2}{\sigma_i{}^2} - \sum_{i=1}^{m}\underline{b}_i\underline{v}_i - \sum_{j=1}^{L}\sum_{i=1}^{S_j}\underline{c}_i{}^{(j)}\underline{h}_i{}^{(j)} - \sum_{i=1}^{m}\sum_{j=1}^{S_1}\frac{\underline{v}_i^T w_{ij}\underline{h}_j^{(1)}}{\sigma_i} -$$
$$\sum_{k=1}^{L-1}\sum_{l=k+1}^{L}\sum_{i=1}^{S_k}\sum_{j=1}^{S_l}\underline{h}_i^{(k)}\widetilde{w}_{ij}^{(k,l)}\underline{h}_j^{(l)} - \sum_{i=1}^{n}d_i u_i - \sum_{k=1}^{L}\sum_{i=1}^{S_k}\sum_{j=1}^{n}\underline{h}_i^{(k)}\underline{U}_{ij}^{(k)}u_j.$$

The joint distributions of the model are defined as:

$$P(\overline{v},\mathbb{H},\underline{u}) = \frac{e^{-E(\overline{v},\mathbb{H},\underline{u})}}{\sum_{\overline{v},\mathbb{H},\underline{u}}e^{-E(\overline{v},\mathbb{H},\underline{u})}},\ P(\overline{v},\underline{u}) = \frac{\sum_{\mathbb{H}}e^{-E(\overline{v},\mathbb{H},\underline{u})}}{\sum_{\overline{v},\mathbb{H},\underline{u}}e^{-E(\overline{v},\mathbb{H},\underline{u})}}.$$

We look for the maximum of the logarithmic likelihood

$$L(\theta) = -\sum_{i=1}^{t}\log p(\overline{v}_i,\underline{u}_i),\quad (\overline{v}_i,\underline{u}_i)\epsilon\mathbb{V},$$

where

$$\mathbb{V} = \{(\overline{v}_i,\underline{u}_i) | i = 1,2,\dots,t\}$$

is a set of training pairs.

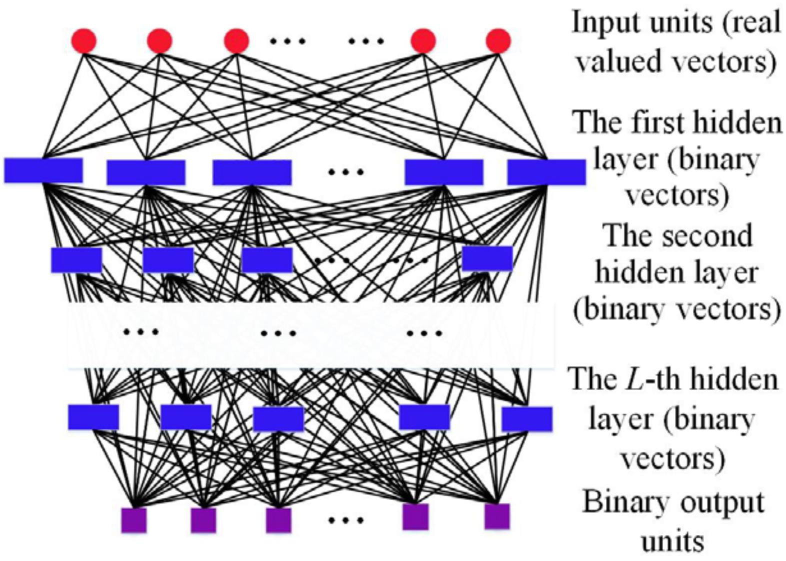

Figure 1 An example for multiconnected Boltzmann machine

We apply the Contrastive Divergence algorithm for training the classification Boltzmann machine. The necessary conditional distributions for the negative phase are defined as follows:

$$P\left(\underline{v}_i\middle|\overline{h}^{(1)}\right) = \mathcal{N}\left(\underline{a}_i + \underline{b}_i + \sigma_i\sum_{j=1}^{S_1}w_{ij}\underline{h}_j^{(1)}, \sigma_i^2 I\right),$$
$$P(u_j|\mathbb{H}) = \mathrm{sigm}\left(d_j + \sum_{k=1}^{L}\sum_{i=1}^{S_K}\underline{h}_i^{(K)}\underline{U}_{ij}^{(K)}\right),$$
$$P(\underline{h}_j^{(1)}|\overline{v},\mathbb{H}_{-1},\underline{u}) = \mathrm{Sigm}\left(\underline{c}_j^{(1)} + \sum_{i=1}^{m}\frac{\underline{v}_i^T w_{ij}}{\sigma_i} + \sum_{l=2}^{L}\sum_{i=1}^{S_l}w_{ji}^{(1,l)}\underline{h}_i^{(l)} + \sum_{i=1}^{n}\underline{U}_{ji}^{(1)}u_i\right),$$

$$P\left(\underline{h}_j^{(l)} \middle| \mathbb{H}_{-l}, \underline{u}\right) := \mathrm{Sigm}\left(\underline{c}_j^{(l)} + \sum_{q=1}^{l-1} \sum_{i=1}^{S_q} \widetilde{w}^{(l,q)} \underline{h}_i^{(q)} + \sum_{q=l+1}^{L} \sum_{i=1}^{S_q} \widetilde{w}_{ji}^{(l,q)} \underline{h}_i^{(q)} + \sum_{i=1}^{n} \underline{U}_{ji}^{(l)} u_i\right), l = 2,3, \dots, L-1,$$

where

$$\mathbb{H}_{-i} = \left(\overline{h}^{(1)}, \overline{h}^{(2)}, \dots, \overline{h}^{(i-1)}, \overline{h}^{(i+1)}, \dots, \overline{h}^{(L)}\right).$$

The first epoch of the Contrastive Divergence algorithm is illustrated on Figure 2. All calculated data are saved in a hard disk. Adaptive learning rate is applied for training of multiconnected Boltzmann machine. The LR (learning rate) search procedure is used with respect to the reconstruction error. In the first epoch we calculate the initial value of the reconstruction error. The choice of the initial value of the variable learning rate is a real challenge. The experiments indicate that the initial guess for the learning rate strongly depends on the input signal i.e. it depends on the input data. We have successful results for the initial value of the learning rate choosing it between 0.01 and 0.001. After the first epoch all information concerning the neural network e.g. hidden units, biases, offsets, weights etc. is saved in the hard disk.

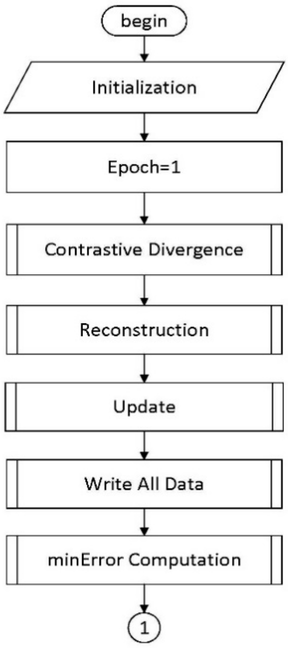

Figure 2 First Epoch of the Contrastive Divergence algorithm with adaptive learning rate

After the second epoch we check whether the reconstruction error is increased or decreased. If the second epoch error is small than the initially calculated error we start the third epoch. In this case the sign of the learning rate is kept and the module of the learning rate is increased. Otherwise, we return to the first epoch loading all saved data from the hard disk. Then we reduce the module of the learning rate and multiply the learning rate by minus one. Imagine that k epochs are already done and the data from the last epoch are stored in the hard disk. We execute the epoch number $k+1$. If the $k+1$-th reconstruction error is smaller than the k-th error we increase the module of the learning rate and continue with the next epoch. In this case the sign of the learning rate should not be changed. If the new obtained error is bigger than the previously calculated one we return to the k-th epoch loading the data from the disk. Additionally, we change the sign of the learning rate and decrease its module. This approach assure monotone decreasing of the reconstruction error. A priory a number ν characterising the smallest acceptable module of the learning rate should be determined. A learning rate close to zero indicates that smaller reconstruction error cannot be found. That is why the inequality

$$|\lambda_k| < \nu$$

defines a stop criterion for training the deep Boltzmann machine. Many neural network developers execute just a fixed number of epochs without satisfying any stop criterion. Even in the case when two adjacent values of the reconstruction error are very close each other there is a real danger to fall into a local minimum or a plateau. In this case the stored weight matrices are used for starting a genetic algorithm [30] to avoid local minima. The genetic algorithm for solving minimization problems needs greater computer resources and much computational work. That is why it could be used in the final stage of the training process. The LR search procedure can be seen in the figure 3. The states of hidden neurons in the positive phase are determined by the mean field approximation. The initial values of the visible and output units in negative phase are taken from the data. The values of the hidden units in the positive phase are initial guesses for the Gibbs sampling algorithm in the negative phase. From the practical point of view we choose $\varepsilon\epsilon(0.02,0.2)$ and $\nu = 10^{-8}$. Anyway the values of ν and ε should be determined empirically from the data.

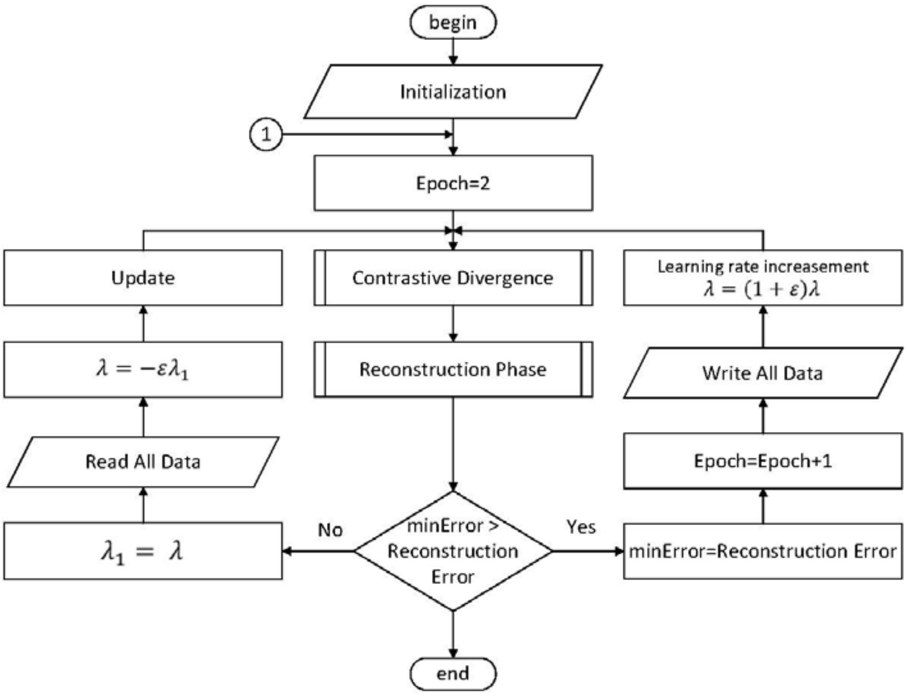

Figure 3 Further epochs of the Contrastive Divergence algorithm

The algorithm for training of the multi connected Boltzmann machine contains positive phase, negative phase, reconstruction, LR search procedure and update. Further the algorithm is described by pseudocode.

```
%Contrastive Divergence algorithm for training
%the classification multi connected L-layerDBM.
%-----------------------------------------------------
%Initialization
%create h̄_i^[0], i = 1,2,...,L randomly
%create u̲^[0], v̄^[0] randomly
%create biases randomly in [0,1]
%create weights randomly in [0,1]
%create offsets
%create other supported data
%create all hidden layers randomly
%Positive phase
%-----------------------------------------------------
for Epoch=1 to Epoch Count do
%Initially the reconstruction error is set to zero
%Min Error= Reconstruction Error
%Contrastive Divergence
begin
  forall(v̄,u̲) ∈ 𝕍 do
```

%i is the current number of a training pair

begin

%The mean field approximation

%The value k indicates the number of iterations

%The positive phase

for $k = 0$ to $K-1$ **do**

begin

%The first hidden layer

for $j = 1$ to s_1 **do**

begin

$$\underline{h}_{1j}^{[K+1]} \coloneqq \mathrm{Sigm}\left(\underline{c}_j^{(1)} + \sum_{i=1}^{m} \frac{\underline{v}_i^T w_{ij}}{\sigma_i} + \sum_{l=2}^{L}\sum_{i=1}^{s_l} \widetilde{w}_{ji}^{(1,l)} \underline{h}_{li}^{[K]} + \sum_{i=1}^{n} \underline{U}_{ji}^{(1)} u_i\right)$$

end j

for $q = 2$ to $L-1$ **do**

begin

%Next hidden layers

for $j = 1$ to s_q **do**

begin

$$h_{qj}^{[K+1]} \coloneqq \mathrm{Sigm}\left(\underline{c}_j^{(q)} + \sum_{l=1}^{q-1}\sum_{i=1}^{s_l} \left(\underline{h}_{li}^{[K+1]}\right)^T \widetilde{w}_{ij}^{(l,i)} + \sum_{l=q+1}^{L}\sum_{i=1}^{s_l} \widetilde{w}_{ji}^{(q,l)} \underline{h}_{li}^{[K]} + \sum_{i=1}^{n} \underline{U}_{ji}^{(q)} u_i\right)$$

end j

end q

end k

end forall

%Approximate values for each unit of each hidden layer

%are obtained after K iterations.

$\widehat{\mathbb{H}} \coloneqq \left(\overline{h}_1^{[k]}, \overline{h}_2^{[k]}, \dots, \overline{h}_L^{[k]}\right)$;

% Negative phase

%---

%Initialization

forall$(\overline{v}, \underline{u}) \in \mathbb{V}$**do**

begin

$\overline{v}^{[0]} := \overline{v}$;

$\underline{u}^{[0]} := \underline{u}$;

end forall;

$\mathbb{H}^{[0]} := \widehat{\mathbb{H}}$;

%The Gibbs sampling iteration

forall$(\overline{v}, \underline{u}) \in \mathbb{V}$ **do**

for $k = 0$ to $K-1$ **do**

begin

%The first hidden layer

for $j = 1$ to s_1 **do**

begin

```
              h_{1j}^{[k+1]} ~ P( h_{1j}^{[k]} | \overline{v}^{[k]}, ℍ_{-1}^{[k]}, u^{[k]} )
          end j
      for l = 2 to L do
      %The next hidden layers
        begin
          for j = 1 to s_l do
            begin
              h_{lj}^{[k+1]} ~ P( h_{lj}^{[k]} | ℍ_{-l}^{[k]}, u^{[k]} )
            end j
          end l
          for j = 1 to m do
          %The visible layer
            begin
              v_j^{[k+1]} ~ P( v_j^{[k]} | \overline{h}_1^{[k+1]} )
            end j
          for j = 1 to n do
          % The output units
            begin
              u_j^{[k+1]} ~ P( v_j^{[k]} | \overline{h}_1^{[k+1]} )
            end j
      end k
end forall
% Approximate values for each unit of each hidden layer
% are obtained after K iterations
\tilde{ℍ} := ( \overline{h}_1^{[k]}, \overline{h}_2^{[k]}, ..., \overline{h}_L^{[k]} );
\tilde{v} := \overline{v}^{[k]};  \tilde{u} := u^{[k]};
%Update
--------------------------------------------------------
for θ ∈ Θ do
  begin
      θ := θ − λ( ∂/∂θ E(\overline{v}, \hat{ℍ}, u) − ∂/∂θ E(\tilde{v}, \tilde{ℍ}, \tilde{u}) )
  end θ
Reconstruction
calculate Reconstruction Error;
if Epoch>1then
  begin
    if Min Error>Reconstruction Error then
      begin
            MinError:=ReconstructionError;
            UpDate;
            Save All Data
            %In this case we have success, the learning
            % rate is increased;
            λ := (1 + ε)λ;
      end
      else begin
           λ_1 := λ;
           load AllSavedData
           λ := −ελ_1;
           removeλ_1;
```

```
          UpDate;
        end
end
else begin
       UpDate;
       Save All Data;
     end
end Epoch
```

III. TESTS

Most of the classifiers are used in various area, from medical domain to contemporary methods in economics. In this paper we focus on the voice remote control. Each voice control system involves a speech recognition tool. Usually the input command is separated into its constituent words. The isolated word is subdivided into m frames. The training of a Boltzmann machine by contemporary computers needs long computational time. Future epochs of computer generation will make applications of the deep neural networks more realistic. We apply a Boltzmann machine for speech recognition with three hidden layers. Its visible layer consists of m units. The frames are cepstral coefficients vectors which contain the most useful features characterizing the input signal. The visible layer is presented as a set of real valued-vectors. The hidden layers consists of binary vectors. The hidden units in a fixed layer have the same length. But the lengths of the hidden units may vary throughout the layers. The output layer is a set of binary numbers. The sum of all units in the output layer is equal to one. More detail description of the model for remote control of moving objects can be found in [28, 29]. The impact of the present paper is on the application of the variable learning rate for training the new multiconnected Boltzmann machine.

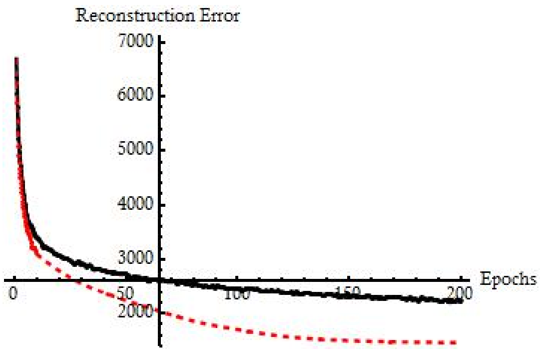

Figure 4 The graph of reconstruction error. The dashed line indicates the application of Algorithm 1 with adaptive variable learning rate, the solid line shows training of the multi connected Boltzmann machine by the same algorithm and a constant learning rate.

We tested the multiconnected Boltzmann machine on an author's experimental data base. The visible layer contains 20 frames, the first hidden layer has 20 hidden units, the second and third hidden layers have 40 and 30 hidden units correspondingly. The output layer has 12 output units. Our goal is to compare the algorithm with LR search procedure and the algorithm with a constant learning rate, Figure 4. The monotone decreasing of the reconstruction error is established, see Figure 4. To present more comprehensively the behavior of the reconstruction error we present its graph between 50-th and 80-th epochs, Figures 5 and 6. Figure 5 indicates a non-monotone decreasing of the error in the case of a constant learning rate equal to 0.00125. The LR search procedure generates better results than the algorithm with a constant learning rate, in Figure 4 even in the case when the genetic algorithm for avoiding of local minima is not used.

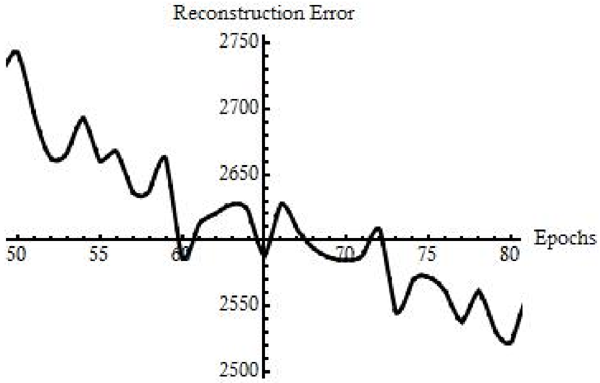

Figure 5 The graph of reconstruction error with a constant learning rate between 50-th and 80-th epoch.

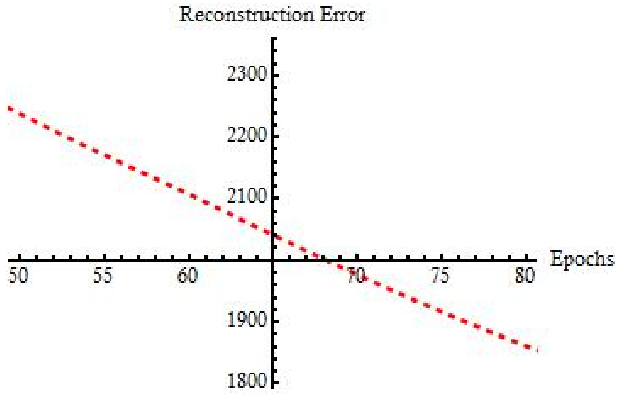

Figure 6 The graph of reconstruction error in the case of the adaptive learning rate between 50-th and 80-th epoch.

IV. CONCLUSION

A new deep multiconnected Boltzmann machine has been developed and studied. An adaptive learning rate is used for training the new Boltzmann machine. LR search procedure is included in the Contrastive Divergence algorithm in order to improve the classification process. Mean field approximation is used in the positive phase to establish the status of hidden neurons. A training algorithm assuring monotone decreasing of the reconstruction error is obtained. The application of the deep Boltzmann machine essentially reduce the number of the hidden units and hence the necessary random access memory. The presented Boltzmann machine is very appropriate for a classifier in voice control systems which requires a high level of accuracy.

REFERENCES

[1] Cynthia Breazeal, "Emotion and sociable humanoid robots," Human-Computer Studies, vol. 59, pp. 119-155, 2003.

[2] S. Wang, M. He, Z. Gao, S. He, Q. Ji, "Emotion recognition from thermal infrared images using deep Boltzmann machine," Comput. Sci. (2014) 8: 609. .

[3] J. Schlüter, C. Osendorfer, "Music Similarity Estimation with the Mean-Covariance Restricted Boltzmann Machine," Proceedings of the Tenth International Conference on Machine Learning and Applications (ICMLA 2011), 2011.

[4] P. Mamoshina, A. Vieira, E. Putin, A. Zhavoronkov, "Applications of Deep Learning in Biomedicine," Artificial Intelligence Research, 2016.

[5] G. Montavon, M. L. Braun, K. Muller, "Deep Boltzmann Machines as Feed-Forward Hierarchies," Proceedings of the 15th International Conference, vol. XX, pp. 798-804, 2012.

[6] P. Xie, Y. Deng, E. Xing, "Diversifying Restricted Boltzmann Machine for Document Modeling," Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1315-1324, 2015.

[7] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," the Journal of machine Learning, 2003.

[8] Z. Cao, S. Li, Y. Liu, W. Li, and H. Ji, "A novel neural topic model and its supervised extension," Conference on Artificial Intelligence, 2015.

[9] Z. Chen and B. Liu, "Mining topics in documents: standing on the shoulders of big data," Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014.

[10] M. Denil, A. Demiraj, N. Kalchbrenner, P. Blunsom and N. de Freitas, "Modelling, visualising and summarising documents with a single convolutional neural network," arXiv preprint arXiv:1406.3830, 2014.

[11] R. Guha, "Towards a model theory for distributed representations," arXiv preprint arXiv:1410.5859, 2014.

[12] G. E. Hinton and R. R. Salakhutdinov, "Replicated softmax: an undirected topic model," Advances in neural information processing systems, 2009.
[13] T. Hofmann, "Probabilistic latent semantic analysis," Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence, 1999.
[14] C. Huang, X. Qiu, and X. Huang, "Text classification with document embeddings," Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, 2014.
[15] H. Larochelle and S. Lauly, "A neural autoregressive topic model," Advances in Neural Information Processing Systems, 2012.
[16] Q. V. Le and T. Mikolov, "Distributed representations of sentences and documents," International Conference on Machine Learning , 2014.
[17] R. Socher, C. C. Lin, C. Manning, and A. Y. Ng., "Parsing natural scenes and natural language with recursive neural networks," Proceedings of the 28th international conference on machine learning, 2011.
[18] N. Srivastava, R. Salakhutdinov, and G. Hinton, "Modeling documents with a deep boltzmann machine," Uncertainty in Artificial Intelligence , 2013.
[19] J. Tomczak , "Application of Classification Restricted Boltzmann Machine with discriminative and sparse learning to medical domains," World Applied Sciences Journal 31 (Applied Research in Science, Engineering and Management), pp. 69-75, 2014.
[20] P. Donovan, "Static Gesture Recognition with Restricted Boltzmann Machines," University of Toronto, Canada, 2007.
[21] Hinton, G. E, Osindero, S., and Teh, Y., "A fast learning algorithm for deep Belief nets," Neural Computation , vol. 18, pp. 1527-1554, 2006.
[22] R. Salakhutdinov, H. Larochelle, "Efficient Learning of Deep Boltzmann Machines," Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), vol. 9, pp. 693-700, 2010.
[23] R. Salakhutdinov, G. Hinton, "Deep Boltzmann Machines," Proceedings of the 12th International Conference on Artificial Intelligence and Statistics (AISTATS), vol. 5, pp. 448-455, 2009.
[24] T.S. Lee, D. Mumford, R. Romero, V. Lamme, "The role of the primary visual cortex in higher level vision," Vision research, vol. 38, pp. 2429-2454, 1998.
[25] A. Chadha, S. Mallik, R. Johar, "Comparative Study and Optimization of Feature Extraction Techniques for Content based Image Retrieval," International Journal of Computer Applications (0975 – 8887), vol. 52, no. 20, pp. 35-42, 2012.
[26] X. Zheng, Z. Wu, H. Meng, W. Li, L. Cai, "Feature Learning with Gaussian Restricted Boltzmann Machine for Robust Speech Recognition," vol. arXiv:1309.6176.
[27] S. Rifai, P. Vincent, X. Muller, X. Glorot, Y. Bengio, "Contractive Auto-Encoders: Explicit Invariance During Feature Extraction," Proceedings of the 28 th International Conference on Machine Learning, 2011.
[28] R.T.Ilarionov, C. Jayne, T. D. Todorov, G. S. Tsanev, S. Y. Yordanov, "A Stochastic Model of Voice Control System," International Journal of Science and Advanced Technology, vol. 5, no. 9, pp. 7-16, 2015.
[29] T. Todorov, G. Tsanev, "Two Hidden Layer Multiconnected Boltzmann Machine for Controlling Moving Objects," British Journal of Mathematics and Computer Science, vol. 19, pp. 7-17, 2016.
[30] K. Liu, L. M. Zhang, Y. W. Sun, " Deep Boltzmann Machines Aided Design Based on Genetic Algorithms," Applied Mechanics and Materials, Vols. 568-570, pp. 848-851, , 2014 .
[31] R. T. Ilarionov, T. D. Todorov, G. S. Tsanev, S. Y. Yordanov, "Signal processing techniques for speech recognition," vol. 50, pp. 44-49, 2014.
[32] H. Larochelle, M. Mandel, R. Pascanu, Y. Bengio, "Learning algorithms for the classification restricted Boltzmann machine," Journal of machine learning research, vol. 13, pp. 643-669, 2013.
[33] A. Courville, J. Bergstra, Y. Bengio, "A spike and slab Restricted Boltzmann Machine," Proceedings of the 14th International Conference on Artificial Intelligence and Statistics, pp. 233-241, 2011.
[34] A. Fischer and C. Igel, "An Introduction to Restricted Boltzmann Machines," CIARP 2012, pp. 14-16.