Modified Deep Belief Network Model and Its Application

Peng Yin
College of Mechanics and Electronics, Shandong Management University, Jinan, China
*Corresponding author’s email: 243609099@qq.com

Abstract. Aiming at the problem that the traditional deep belief network (DBN) model has long training time and it is difficult to find the global optimum because of the fixed step. A modified DBN model was proposed. In the DBN pre-training phase, the adaptive step (AS) was introduced into the Contrastive Divergence (CD) algorithm to solve the problem of finding the global optimum difficulty because of the fixed step, improving the accuracy and oscillation resistance. In the fine-tuning phase, the BFGS quasi-newton method (QNM) was used to speed up the convergence and reduced the training time. The modified DBN model was applied to the fuel consumption prediction of the descending section of the aircraft. The experimental outcomes prove that compared with the traditional DBN and its variant model, the modified DBN model improves the convergence speed and the prediction accuracy.

Keywords. Deep belief network, Adaptive step, BFGS quasi-newton method, Fuel consumption prediction

1. Introduction
Deep Belief Network (DBN) contains multiple hidden layers as a deep learning model. It has strong learning ability and complex function relationship expression ability. It has been widely used in nonlinear system prediction in recent years [1,2]. Zhang Pengcheng et al. [3] proposed a DBN-based precipitation prediction model to achieve accurate precipitation prediction. Li Zichen et al. [4] studied the activity structure data of PD pathogenic genes and proposed a DBN-based PD pathogenic gene activity prediction method, which improved the prediction accuracy and stability. Zhou Wenjie et al. [5] proposed a DBN-based user complaint prediction model, which has improved prediction accuracy compared with previous user complaint models.

There are still many problems in the theory and learning algorithm, such as slow convergence speed and difficulty in finding global optimality by fixed steps, even though DBN is widely applied in the domain of nonlinear system prediction [6]. There are two stages that the DBN model training process comprised, which are pre-training and fine-tuning. In the pre-training process, a fixed step size will cause the contrast divergence (CD) algorithm to appear “premature” or oscillations, which makes it tough to find the global optimum problem. In the fine-tuning process, when using the Gradient Descent to adjust the parameters, the closer to the target value, the slower the convergence speed, resulting in a long training time.

In order to solve these problems, an improved DBN model is proposed, which introduces adaptive step size (AS) into the CD algorithm in the pre-training stage, and continuously adjusts the step size according to the parameter update direction after two consecutive iterations, thereby avoiding the
problem of finding the global optimum caused by the fixed step size is solved. In the fine-tuning stage, the BFGS quasi-Newton method is carried over into accelerate network learning, which solves the problem of slow model convergence and long training time. In this paper, the improved DBN model is applied to the prediction of fuel consumption during the aircraft descent phase. The results show that the improved DBN model can enhance the convergence rate and forecast precision.

2. Improved DBN model

2.1 Improved DBN model structure
Hinton et al. proposed the concept of deep learning in 2006 [7]. The DBN is an extensive used network model in deep learning. It is on the basis of the energy probability distribution and consists of multiple restricted Boltzmann machine (RBM) models and BP neural network. Similar to the traditional DBN, the improved DBN model structure is shown in Figure 1.

![Figure 1. Improved DBN model architecture](image)

2.2 Introducing the pre-training process of adaptive step size
There are two stages that the DBN model training process comprised, which are pre-training and fine-tuning. In the pre-training stage, the CD algorithm is carried over into perform unsupervised training layer by layer for each RBM [8], to initialize the network parameters.

RBM [9] is a two-layer undirected graph model, including a visible layer $H$ and a hidden layer $V$, as shown in Figure 2. The neurons in the visible layer and the hidden layer are not connected in the layer, and between the layers are completely connected. The visible layer is to receive input data; the hidden layer is to reconstruct the visible layer and can be regarded as a feature extraction layer.

![Figure 2. RBM structure](image)

Presume that they are binary distributions included all hidden layer units $h$ and visible layer units $v$, namely $\forall i,j, v_i \in \{0,1\}, h_j \in \{0,1\}$. For a group of preseted conditions $(v, h)$, the energy function of RBM
The energy function, the joint distribution of this group of conditions \((v, h)\) can be obtained as

\[
P_\theta(v, h) = \frac{e^{-E_\theta(v, h)}}{Z_\theta}
\]

\[
Z_\theta = \sum_{v, h} e^{-E_\theta(v, h)}
\]

Among them, \(m\) and \(n\) are the number of the visible layer element and the hidden layer element respectively; \(\theta = \{w_{ij}, a_i, b_j\}\) is the RBM parameter; \(w_{ij}\) is the connection weight between the hidden layer element and the visible layer element; \(a_i\) and \(b_j\), respectively represent the offset of visible layer element and hidden layer element. According to the energy function, the joint distribution of this group of conditions \((v, h)\) can be obtained as

\[
P_\theta(v, h) = \frac{e^{-E_\theta(v, h)}}{Z_\theta}
\]

\[
Z_\theta = \sum_{v, h} e^{-E_\theta(v, h)}
\]

\(Z_\theta\) is the normalization factor. According to the structural properties of RBM, on the basis of a given visible layer unit \(v\) (or hidden layer \(h\), the state of each hidden layer unit \(h\) (or visible layer unit \(v\)) is independent of each other, so the \(j\)-th hidden layer is calculated. The conditional probability distributions of the layer unit and the \(i\)-th visible layer unit are respectively

\[
P_\theta(h_j = 1|v) = \text{sigmoid}(b_j + \sum_i v_i w_{ij})
\]

\[
P_\theta(v_i = 1|h) = \text{sigmoid}(a_i + \sum_j h_j w_{ij})
\]

Among them,

\[
\text{sigmoid}(x) = \frac{1}{1 + \exp(-x)}
\]

When a set of training sample set \(S=\{v^1, v^2, \ldots, v^n\}\) is given, the goal is make the likelihood function attain to maximum, just as follows:

\[
L_{\theta, S} = \sum_{n=1}^{N} \log P(v^n)
\]

The quantity of training samples is expressed by \(N\). For the sake of obtaining the optimal parameter \(\theta\), the gradient ascent method is to attained to the maximum value of \(L_{\theta, S}\). The update formula of parameter \(\theta\) is:

\[
\theta_t = \theta_{t-1} + \eta \frac{\partial \ln L_{\theta, S}}{\partial \theta}
\]

Among them, \(\eta\) represents the step size. Assuming there is only one training sample, the partial derivatives of the log-likelihood function \(L_\theta\) for \(w_{ij}, a_i\) and \(b_j\) in the parameter \(\theta\) are respectively.

\[
\frac{\partial L_\theta}{\partial w_{ij}} = (v_i h_j)^{\text{data}} - (v_i h_j)^{\text{model}}
\]

\[
\frac{\partial L_\theta}{\partial a_i} = (v_i h_j)^{\text{data}} - (v_i h_j)^{\text{model}}
\]

\[
\frac{\partial L_\theta}{\partial b_{ij}} = (h_j v_i)^{\text{data}} - (h_j v_i)^{\text{model}}
\]
\[ \{ \}_{\text{data}} \] and \[ \{ \}_{\text{model}} \] respectively represent the mathematical expectation of the two probability distributions of \( P_{\theta}(h|v) \) and \( P_{\theta}(v,h) \). Since \[ \{ \}_{\text{model}} \] is difficult to obtain in actual calculations, the K-step Gibbs sampling method in the CD algorithm is used to obtain the approximate values [10], its approximate value is expressed as \( \{ \}_{k} \).

Since the multiple iterations are required in each RBM in the CD algorithm, and the direction of parameter update is different after each iteration. It is necessary to select an appropriate step size \( \eta \). If the selected \( \eta \) is large, it is easy to cause oscillation; if \( \eta \) is small, the convergence speed will be slow. To solve this problem, an adaptive step size (AS) method is proposed [11].

In the RBM training process, we can obtain a parameter renew direction after two successions iterations, which is the basis of the as method adjusts \( \eta \), and \( \frac{\partial \ln L_{\theta,t}}{\partial \theta} \) is abbreviated as \( \Delta \theta_{t} \) then

\[
\eta = \begin{cases} 
\alpha \eta', & \Delta \theta_{t} \cdot \Delta \theta_{t-1} > 0 \\
\beta \eta', & \Delta \theta_{t} \cdot \Delta \theta_{t-1} < 0 \end{cases}
\] (12)

Among them, “\( \alpha>1 \)” indicates the increase coefficient of \( \eta \), “\( \beta<1 \)” indicates the decrease coefficient of \( \eta \), and \( 0<\beta<\alpha \). If the two consecutive parameter update directions are the same, \( \eta \) increases. If the directions are opposite, \( \eta \) decreases. After the introduction of the adaptive step size, the parameter update mechanism has changed. The update formulas for the parameters \( w_{ij}, a_{i} \) and \( b_{j} \) are respectively

\[
w_{ij} = w_{ij} + \eta^{*} \left( \langle v_{i} \mid \bar{j} \rangle_{\text{data}} - \langle v_{i} \mid \bar{j} \rangle_{k} \right)
\] (13)

\[
a_{i} = a_{i} + \eta^{*} \left( \langle v_{i} \rangle_{\text{data}} - \langle v_{i} \rangle_{k} \right)
\] (14)

\[
b_{j} = b_{j} + \eta^{*} \left( \langle \bar{i} \rangle_{\text{data}} - \langle \bar{i} \rangle_{k} \right)
\] (15)

Among them, \( \eta^{*} \) represents the adaptive step size. Take the connection weight \( w_{ij} \) as an example to illustrate the effect of introducing an adaptive step size. When the direction of this update \( \frac{\partial L_{\theta}}{\partial w_{ij}} \) is the same as the previous time, the step size \( \eta^{*} \) increases, and the adjustment range of \( w_{ij} \) is larger, which speeds up the adjustment speed during stable adjustment; when the update direction has a different sign, it indicates that there is some instability and the step size \( \eta^{*} \) decreases. Make the adjustment range of \( w_{ij} \) smaller and stabilize. Therefore, lead the Adaptive-Step Algorithm into the CD algorithm during the pre-training stage can accurately and quickly find the optimal solution that satisfies the objective function.

### 2.3 Fine-tuning process based on BFGS quasi-Newton method

In the process of DBN fine-tuning, we usually adopt BP algorithm supervised parameter of fine-tuning, and based on the steepest descent method, the parameters are adjusted trace the negative gradient direction of the goal function. However, the closer this approach is to the target value, the slower the convergence speed, which leads to the problem of long training time. Therefore, this article uses BFGS quasi-Newton method [12] to replace the steepest descent method to speed up the learning speed.

The BFGS quasi-Newton method is an improvement to the Newton method. The objective function value and the first derivative can be utilized to construct an approximate Hessian matrix, which solves the problem of the need to calculate the Hessian matrix in the Newton method, which causes a large workload. At the same time, it retains the advantage of fast convergence which the Newton's method has.

Based on the above method, it can obtain an objective function \( F \) as following formula:

\[
F(\gamma) = \frac{1}{n} \sum_{k=1}^{n} (\bar{T} - Y)^{2}
\] (16)
Among them, $\hat{Y}$ represents the DBN target output, that is, the true value of fuel consumption in aircraft descent stage; $Y$ represents the DBN predicted output, that is, the predicted value of the fuel consumption output by the model; $\hat{Y} - Y$ represents the output error. In the fine-tuning stage, the weight of DBN $\gamma$ is adjusted according to equation (17).

$$\gamma_{k+1} = \gamma_k + \alpha_k d_k$$  \hspace{1cm} (17)

Among them, $d_k$ represents the search orientation of the k-th iteration; $\alpha_k$ represents the step size of the k-th search orientation. Generally, a one-dimensional search is performed along the search direction to obtain the optimal step size.

$$\alpha_k = \arg \min_{\alpha} f(\gamma_k + \alpha d_k)$$  \hspace{1cm} (18)

Let the gradient of the goal function $F$ be $g$, namely

$$g = \nabla F(\gamma)$$  \hspace{1cm} (19)

Let the search orientation $d_k$ of the k-th iteration be

$$d_k = -B_k^{-1} g_k$$  \hspace{1cm} (20)

Among them, $B_k$ is the approximate Hessian matrix.

$$B_{k+1} = B_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{g_k g_k^T}{s_k^T B_k s_k}$$  \hspace{1cm} (21)

Among them,

$$s_k = \gamma_{k+1} - \gamma_k$$  \hspace{1cm} (22)

$$y_k = g_{k+1} - g_k$$  \hspace{1cm} (23)

Therefore, drawing the adaptive step-size (AS) into the pre-training stage can solve the problem that fixed step-size is not easily find the global optimum. During the fine-tuning stage, drawing the BFGS quasi-Newton method into accelerating the network learning speed and reducing the training time. The specific algorithm flow is shown in Figure 3.
3. Improved the application and verification of the DBN model

This research draws into the improved DBN model to predict the fuel consumption during the aircraft descent phase and verify the performance of the model. The fuel consumption during the aircraft descent phase refers to the amount of fuel consumed by the aircraft during the descent phase. Along with the constant improvement of air transportation, the consumption of aviation kerosene is increasing year by year. The civil aviation industry is facing increasing challenges in terms of energy and environment. For the industry, accurately predicting fuel consumption and implementing refined fuel management can help reduce fuel consumption. Therefore, a well-designed fuel consumption model is of great significance for improving fuel utilization and reducing civil aviation carbon emissions [13,14].

3.1 Experimental sample set

The fuel consumption during the aircraft descent phase is affected by many factors. The fuel consumption and many factors affecting fuel consumption during the descent phase are recorded in the Quick Access Recorder (QAR), including aircraft weight, descent distance, descent rate, wind speed and many other parameters. The degree of influence of each influencing factor on fuel consumption is not the same, and there is a non-linear relationship between each influencing factor and fuel consumption. After analyzing the degree of influence of each influencing factor on fuel consumption, 10 parameters including total aircraft weight, descent distance, descent rate, wind speed, wind direction, total atmospheric temperature, pitch angle, angle of attack, left engine speed, and right engine speed are finally selected as the
importation named $X$ of fuel consumption model. And the fuel consumption value of the descending section is used as the model output $Y$ to establish a prediction model. The experimental records are derived from the QAR records of the descent section of an airline on the same route and aircraft type. A total of more than 8,000 sets of descent data are obtained. Stochastic extract 80% of the data as the training sample, and the rest 20% as the test sample.

Normalize the input data into the same dimension. The normalization formula is as follows

$$x_k = \frac{x-x_{\min}}{x_{\max}-x_{\min}}$$

(24)

Among them, $x$ represents the input sample value of fuel consumption factors, $x_k$ represents the normalized value of the sample, and $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values of the sample, respectively.

3.2 Model evaluation indicators

This paper evaluates the performance of the model from two aspects: prediction accuracy and time complexity. Among them, the prediction accuracy evaluation index includes RMSE, MAPE and R, and the time complexity evaluation index is operation time (T). The indicators are defined as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}$$

(25)

$$MAPE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$

(26)

Among them, $y_i$ is the estimated value of fuel consumption in the descending stage, $\hat{y}_i$ is the actual value, and $\bar{y}$ is the estimated average value.

3.3 Experimental setup

Based on the reconstruction error-based DBN network depth determination method proposed by Pan Guangyuan et al. [15], the quantity of hidden layers of the improved DBN model is determined to be 3, and choose 10 fuel consumption as the model input, so the quantity of visible layer nodes is 10. Each hidden layer and BP network respectively have 20 nodes and 5 nodes.

The compiling software and computer operating environment of this simulation experiment are as follows: the compiling software is MATLAB 2009a version, the computer processor is Intel-I7 CPU, and the memory is 4GB.

3.4 Experimental results and analysis

For the sake of testing the characteristic of the improved DBN model, the improved DBN model was compared with the traditional DBN, AS-DBN and BFGS-DBN models. Figure 4 shows the partial fuel consumption prediction outcomes of the improved DBN model. Figure 5 displays the relative error between the partial fuel consumption predictions of the improved DBN model and the true value. Table 1 details the prediction precision and time complexity of the improved DBN and the other three models. The experimental result data.
According to Figures 4 and 5, it can be seen that the relative deviation between the forecasted value of fuel consumption of the improved DBN model and the true value is mostly within the interval [-0.04, 0.04]. Considering that fuel consumption during the descent section will also be affected by weather conditions and route. Due to the influence of other factors such as congestion, the allowable error exists, and it is feasible to improve the DBN model within the allowable range of error.

Table 1 details the AS-DBN model with adaptive step size has higher prediction accuracy than the traditional DBN model, and the BFGS-DBN model with BFGS quasi-Newton method reduces the time.
complexity compared with the traditional DBN model; compare with DBN, AS-DBN and BFGS-DBN three models, the improved DBN model has significant improvements in both prediction accuracy and time complexity, and has a greater advantage in fuel consumption estimation during the descent phase of the aircraft.

4. Conclusion
This article improves the traditional DBN from the two stages of pre-training and fine-tuning. In the pre-training stage, the adaptive step size (AS) is introduced to avoid the problem of finding the global optimum caused by the fixed step size, and improve the prediction accuracy and anti-vibration ability; in the fine-tuning stage, drawing into the BFGS quasi-Newton method to replace the steepest decline Method, speed up the network learning speed. After the above improvements, utilizing the new model to predict fuel consumption during the descent section of the aircraft. Experimental outcomes show that, the improved DBN model proposed in this paper has been significantly improved for forecast accuracy and time complexity, thus verifying its effectiveness and feasibility in the application of nonlinear system prediction.

References
[1] Rosa E D L, Yu W. 2016 Randomized algorithms for nonlinear system identification with deep learning modification vol 364 no 365 (Information Sciences) pp 197-212.
[2] Lecun Y, Bengio Y, Hinton G. 2015 Deep learning vol 521, no 7553 (Nature) pp 436-444.
[3] Zhang P C, Zhang L, Wang J M. 2017 A model for precipitation for easting of multiple environmental factors based on deep network vol 34, no 9 (Computer Applications and Software) pp 240-245.
[4] Li Z C, Tian S W, Liu J Y, et al. 2017 Prediction of PD disease gene activity based on deep learning vol 34 no 9 (Computer Applications and Software) pp 183-187.
[5] Zhou W J, Yan J F, Yang L. 2017 Research on prediction model of complaint based on deep learning vol 34 no 5 (Application Research of Computers) pp 1428-1432.
[6] Liu J W, Liu Y, Luo X L. 2014 Research and development on deep learning vol 31 (Application Research of Computers) pp 1921-1930.
[7] Zhang J Y, Wang H L, Guo Y, et al. 2018 Review of deep learning vol 35 no 7 (Application Research of Computers) pp 1921-1928.
[8] Hinton G E, Osindero S, Teh Y. 2006 A fast learning algorithm for deep belief nets vol 18 no 7 (Neural Computation) pp 1527-1554.
[9] Zhang X, Shi L, Shang B, et al. 2018 Research on spark parallelization of deep confidence network in microblogging emotion classification vol 35 no 2 (Computer Applications and Software) pp 48-53.
[10] Fischer A, Igel C, 2014 Training restricted Boltzmann machines: an introduction vol 47 no 1 (Pattern Recognition) p 25-39.
[11] Lopes N, Ribeiro B, 2014 Towards adaptive learning with improved convergence of deep belief networks on graphics processing units vol 47 no 1 (Pattern Recognition) pp 114-127.
[12] Wu G Z, Liang H Y. 2014 A modified BFGS method and its convergence vol 18 no 11 pp 43-47.
[13] Chen J J, Xiao G P. 2014 Minimum sample size analysis of aircraft segment fuel consumption interval estimation vol 35 (Computer Engineering and Design) pp 4356-4359.
[14] Grote M, Williams I, Preston J. 2014 Direct carbon dioxide emissions from civil aircraft vol 95 no 1 (Atmospheric Environment) pp 214-224.
[15] Pan G Y, Chai W, Qiao J F. 2015 Calculation for depth of deep belief network vol 30 no 2 (Control and Decision) pp 256-260.