

The Fault Detection of Aero-engine Sensor Based on Deep Belief Networks

GUO Chuang , ZHENG Xiao-fei, Yao-bin

College of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an 710038, China;

Keywords: Aeroengine sensor; Fault Detection; Deep learning; Deep belief network; Flight parameter

Abstract. Due to its high complexity and accuracy, the faults of aeroengine sensor can not be described with precise mathematical model and it is hard to detect the faults using traditional redundant method. Focus on this problem, a big data-deep learning based model is established to detect the faults of sensor of the engine air exhaust temperature. The classified model using deep belief network(DBN) is built firstly and is trained using a great number of data collected by flight parameter recorder. The model is able to classify the fault through feature learning layer by layer. The results of simulation experiment show that the accuracy with artificial feature extraction is 98% while 96.6% without it. The accuracy of this model is higher than the model using BP neural net and support vector machine(SVM) in both conditions which shows the superiority of the DBN algorithm in sensors fault diagnosis.

1. Introduction

As the heart of the plane, the aeroengine directly affects the flight safety. In order to assure the safety, it's necessary to monitor the engine precisely and in real time. The effective monitoring is based on the normal operation of sensors. However, the sensors of the engine normally work under the harsh circumstances like high temperature, pressure and speed. Statistically, the sensors faults occupy 80% [1-2] of the amount of engine control system faults. In a word, researches on the fault detection techniques of sensor means greatly for ensuring the flight safety.

The common detection techniques include hardware redundancy, and methods based on AI. Formerly, the fault detection of sensors relies on hardware redundancy which requires the staff have the profound knowledge of the structure and the theory of the sensor. At the same time, the better effects we demand, the more precise mathematical model needs to be built, which is feeble when meeting with more and more nonlinear systems. Therefore, the paper mainly researches on how to detect the faults through the technique based on artificial intelligence. Deep belief is an emerging research hotspot in machine learning, it stimulates the abstract processes of human brain which learns from the training samples and acquires the statistical law so as to solve the classification problem. Compared with the traditional neural net, deep belief can avoid the defects like over fitting, local minimum, and gradient descent. It adopted the unsupervised greedy layer-wise learning algorithm to train the network which could not only escape from the gradient descent but also continuously abstract new features from high-dimensional large scale data. This method effectively avoids the tedious processes of abstraction by hand, saves lots of manpower, and increases the accuracy of classification.

The paper takes the fault detection of the engine exhaust air temperature sensor as an example, and comes up with fault detection method based on the large scale data and deep belief. Because of the widely application of the flight data recorder, plenty of high dimensional large scale data has been recorded. There is a high correlation among them, these data meets some kind of relationship when the sensor works normally, while this relationship breaks when sensor failure happens. Therefore, we can train the deep belief network by the data sets, and realize the classified function.

2. The Collection And Pretreatment of Information.

DBN is the method been most widely researched and applied in deep belief theory, it shows the outstanding classified effects in reference[3-7], which is the exact method adopted in the paper. In order to realize the intelligent detection of deep belief, a great amount of data is needed. The widely use of flight data recorder accumulates quantities of data. Pick up 7000 sets of data at random from a flight parameter recorder of some type of a domestically-made airplane. The data includes all the parameters in table 1, and build a 13-dimension data set. Due to the inevitable situation of flight data corruption, pretreatment needs to be done. The paper adopted the method based on the state matching and support vector machine(SVM) to make up the missing flight data. The output layer of DBN is the joint probability distribution of samples, and the output value is between 0 and 1, thus, the input data needs to be normalized on the [0,1] interval. After all of the pretreatments are finished, the data base is consist of 7000 13-dimension flight parameter data.

$$\overline{X}_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

The equation above: X_i is the sample data; X_{\max} , X_{\min} are the minimum and the maximum of sample data; \overline{X}_i is the sample data after normalization.

Tab.1 Thirteen dimensional data used in this paper

The position of low pressure guide vane	α_1	rad
The position of high pressure guide vane	α_2	rad
low - pressure rotor speed	N_1	r / min
high - pressure rotor speed	N_2	r / min
Nozzle diameter	d	mm
Throttle angle	φ	rad
Oil pressure	p	Pa
indicated airspeed	v	m / s
engine air exhaust temperature	T_{EG}	K
The position of swash plate of aircraft inlet	θ	rad
Height of ADS	h	m
Fuel left	FF	%
Flight time	t	s

3. The Fault Detection Based on DBN.

Modeling according to the theory of the DBN, the specific processes are as follow:

3.1 The Partition of The Data.

Divide the 7000 sets of data into 6850 training data $\{(\alpha_i, \alpha_2, N_1, N_2, d_i, \varphi_i, p_i, v_i, T_{EG_i}, \theta_i, h_i, FF_i, t_i), Y_i\}$ and 150 test data $\{(\alpha_j, \alpha_2, N_1, N_2, d_j, \varphi_j, p_j, v_j, T_{EG_j}, \theta_j, h_j, FF_j, t_j), Y_j\}$, and Y_i 、 Y_j are the tags of the corresponding data respectively, which is normal or faulted. [3-7]

3.2 The Construction of DBN.

DBN is consist of a stack with a series of restricted Boltzmann machine(RBM). The basic theory of RBM is from classical thermodynamics[9], which is that the temperature(energy) of the particle is inversely proportional to its existence probability, in other words, the more energy the particle obtains, the lower possibility it exists, otherwise the higher. By the means of 'greedy layer-wise learning', the RBM network turns into a stable state with minimum energy, and acquire high-order abstract features. Pile a series RBM and get a highly complex directed acyclic network consists of several random variables. Figure 1 shows us a deep belief network with 3 hidden layers from which we can see that two layers on the top are undirected models while other layers are directed and the direction is from up to down. The merits is that the output of the lower layers can provide a considerable association for top layers, and binds themselves with their memories, which is flexible

for the fine-tuning [10]. The input of visible layer V is sample, h_1 、 h_2 、 h_3 are the first, second and the third hidden layers; W is the weight of the connections between two neurons nearby. The output of belief network with l hidden layers is:

$$P(v, h^1, h^2 \cdots h^l) = \left(\prod_{k=0}^{l-2} P(h^k | h^{k+1}) \right) P(h^{l-1}, h^l) \quad (2)$$

In the equation $v = h^0$, $P(h^k | h^{k+1})$ is the conditional distribution between the k th RBM visible layers and hidden layers. It shows that the overall network is exactly learning this joint probability distribution.

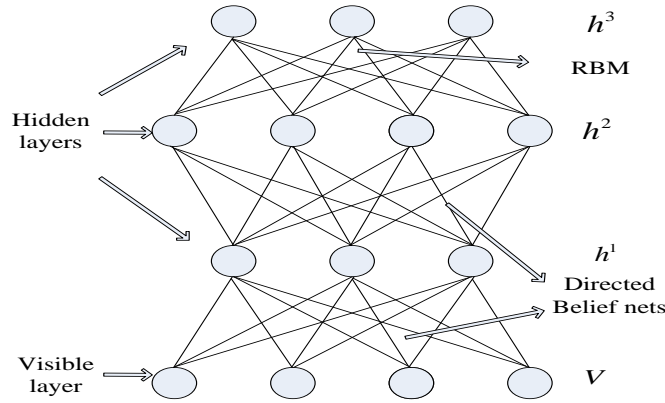


Fig.1 Structure of DBN with three hidden layers

For the know ability of the input data sets, the number of neurons on visible layer v is 13. But the quality of the network is affected by the number of neurons on the hidden layers to a great extent, and there's no exact statement about how to select the number of the neurons. After lots of experiments, we get $n_{h_1} = 100$, $n_{h_2} [15 \sim 35]$ take one from the interval, $n_{h_3} [5 \sim 25]$ take one from the interval. So as to acquire the optimal numbers of nodes, we make use of the reconstruction error to optimize, and make the contour error in figure 2.

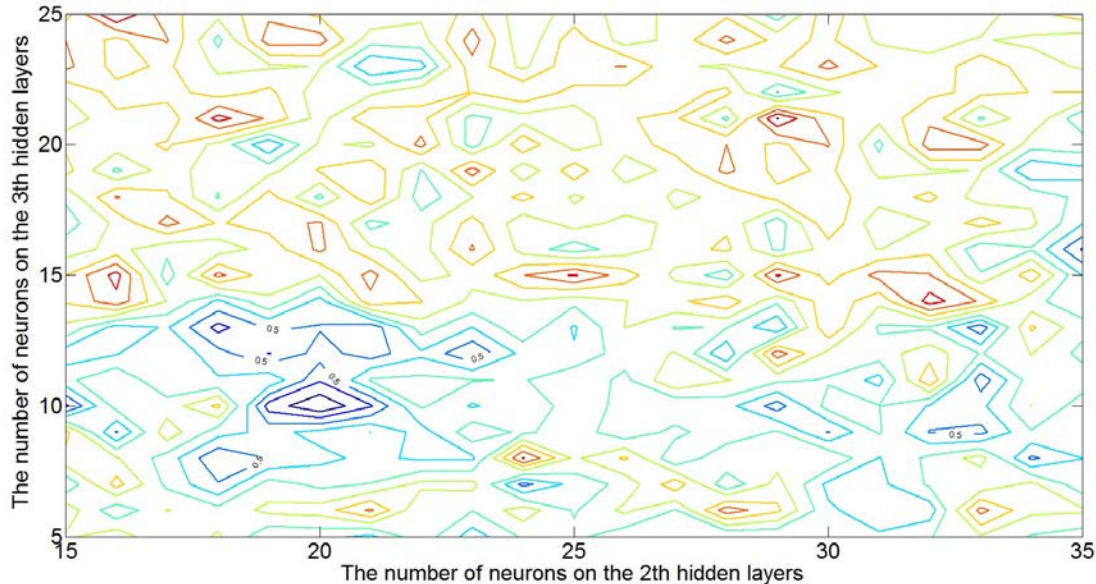


Fig.2 Change of training sample reconstruction error with node number n_{h_2} and n_{h_3}

The specific processes of calculating the reconstruction error are shown as table 2. Seen from figure 2, the reconstruction error is minimum when $n_{h_2} = 20$, $n_{h_3} = 10$. As for the partition problem is a bipartition one(normal, failure), the number of the output layer is 2, the final structure is 13-100-20-10-2.

Tab.2 Steps of calculating reconstruction error

initialization $Error = 0$
Enter the training sample $v^{(i)}$, the number of the hidden unit m of RBM , the maximum cyclic iteration times T ;
For all $v^{(i)}, i \in \{1, 2, \dots, K\}$
Take sample of hidden layer, calculate $P(h v^{(i)}, \theta)$, extract $h \in \{0, 1\}$ from condition distribution;
Take sample of visible layer, calculate $P(v_1^{(i)} h, \theta)$, extract $v_1^{(i)} \in \{0, 1\}$ from condition distribution;
$Error = Error + \ v_1^{(i)} - v^{(i)}\ $
End
Return overall $Error$

3.2 The Training of DBN

The training of the DBN is divided into two main steps. (1) Train each layer of RBM with unsupervised greedy layer-wise learning algorithm. (2) Fine tune the whole network and make it optimal. The specific processes are shown in table 3.

Tab.3 Steps of training DBN

Step1	Take $v = h^0$ as input, train the first RBM, let it get stable
Step2	Take the joint probability distribution obtained from the 1st RBM as the input of the 2nd RBM visible layer, wait until it stays stable.
Step3	Repeat step 2 until the last RBM
Step4	Take the maximum likelihood function as the objective function, and fine tune the parameters in every layer with back propagation algorithm, in order to optimize the network.

In the first 3 steps, we acquired the parameter set $\theta = (w_{ij}, a_i, b_j)$ of the RBM through learning the training data. w_{ij} is the connection weight between the i th node in visible layer and the j th node in hidden layer; a_i is the bias of the i th node; b_j is the bias of the j th node. The parameter adjustment algorithm of RBM is as follow:

$$\Delta w_{ij} = \varepsilon (\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{model}) \quad (3)$$

$$\Delta a_i = \varepsilon (\langle v_i \rangle_{data} - \langle v_i \rangle_{model}) \quad (4)$$

$$\Delta b_j = \varepsilon (\langle h_j \rangle_{data} - \langle h_j \rangle_{model}) \quad (5)$$

In the equations above, it takes exponential time to calculate $\langle \rangle_{model}$, the contrast divergence algorithm can reduce the computation, the new adjustment algorithm is as follow:

$$\Delta w_{ij} = \varepsilon (\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_1) \quad (6)$$

$$\Delta a_i = \varepsilon (\langle v_i \rangle_{data} - \langle v_i \rangle_1) \quad (7)$$

$$\Delta b_j = \varepsilon (\langle h_j \rangle_{data} - \langle h_j \rangle_1) \quad (8)$$

The $\langle \rangle_1$ is the reconstruction sample obtained from the gibbs sampling of samples. ε in the equation above is learning ratio, it shows each time the size of the parameter adjustment, which is usually between 0.005~0.200[11], so we take $\varepsilon = 0.15$. Owing to the local minimum that the learning processes often trapped in, the momentum is introduced, this method can avoid the consistent direction of the parameter update and the gradient, the adjustment method is as follow:

$$w_{ij}^{t+1} = m w_{ij}^t + \varepsilon \frac{\partial \theta}{\partial w_{ij}} \quad (9)$$

In the equation, m is momentum, in the paper we take $m=0.5$; t is the iterations.

When set up the parameters above, the iterations of RBM needs setting, a weight update of RBM equals an iteration, according to plenty of experiments that the variety of reconstruction error is slight when the iterations is more than 250, that's the reason why we set each layer iteration 250. When all the training of the RBM is done, the first 3 steps of DBN training are over.

Do global fine tune in step 4, take original training data as supervised data and conduct supervised learning of the overall network, optimize the deep network with back propagation algorithm, measure the approximation degree of the input and output probability with crossing entropy, the equation goes:

$$H(q, p) = -\sum_{i=1}^n q_i \lg p_i - \sum_{i=1}^n (1 - q_i) \lg (1 - q_i) \quad (10)$$

In the equation: q, p are the probability distributions of input and reconstruction sample respectively.

Seen from the equation(10), the smaller the crossing entropy is, the more complete features of sample it represents. After lots of experiments, the number of fine tuning times in the paper is 150.

For the sake of better analysis of abstraction of DBN features, figure 3 is the abstraction curve of each layer when the data go through DBN training, the ordinate in the figure is reconstruction feature.

Known from the figure 3: (1) the reconstruction character of the 3rd hidden layer where sensors work normally is smoother than sensors with faults, it is because that data in each dimensions meets some relations and the features are more obvious when sensors work normally, on the contrary, the data doesn't satisfy some relations and the features are out of order. (2) the deeper the hidden layer is, the more abstract the features are.

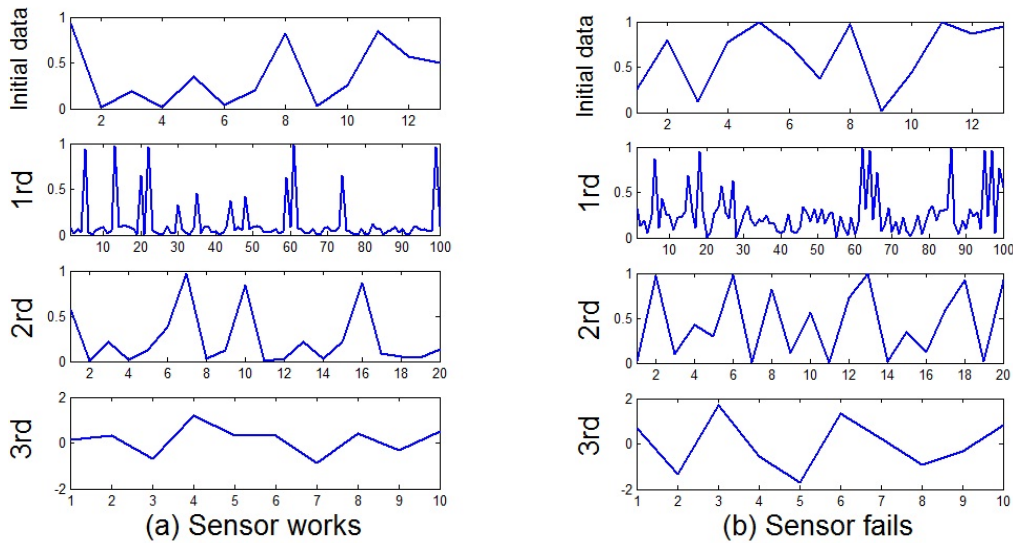


Fig.3 Layer-wise feature extraction curve of data in DBN

4. The Comparison and Analysis of Fault Detection Effect

We can test the sample after the DBN has been trained. The recognition effect may vary because the DBN is a model based on probability. To obtain a better recognition effect, the computing result is the average of 50 times running of the program, the accuracy rate is shown visually through figure 4.

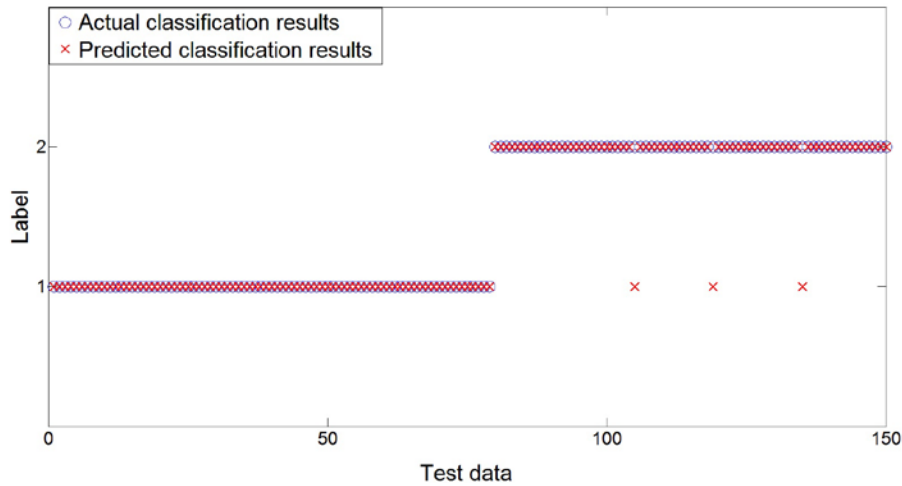


Fig.4 Classification results of test set

In the interest of comparing the effect of fault detection, the paper compared the accuracy of the DBN model and the mainstream detections(SVM and BP nerve net). Meanwhile, in order to show the great capability of abstracting features automatically of deep belief, the paper sets two contrast test groups. Group 1: no manual feature extraction, detect faults with original data. Group 2: apply PCA, relevance analysis, and artificial exhaustion on original data to extract the primary feature and recognize the primary features, and find what affect fault detection are: $N_1, N_2, d, \varphi, T_{EG}$ then detect the fault with intelligent algorithm.

Seen from table 4, the contrast of six fault detection methods, we can conclude that: (1) the classification accuracy of fault detection based on DBN is far higher than others(BP network and SVM) with or without artificial feature extraction. (2) compared with training sets and testing sets of the same method, the accuracy based on DBN is close whereas the other methods turn out differently, which shows that the detection based on DBN will not trap in over-fitting due to the increasing training times. (3) compared with two groups, there's been a sharp rise in the classification accuracy of detections based on BP network and SVM after the artificial feature extraction while the detection based on DBN falls. This fully proves that DBN can automatically extract the features from high-dimension and large scale of data, the higher the dimensions and the larger scale of data is, the more characters the DBN learns and the more precisely it describes the sample. (4) the reason why the accuracy of DBN declines after artificial feature extraction probably is that the 8 dimension data $\alpha_1, \alpha_2, p, v, \theta, h, FF, t$ are eliminated and reduce some hidden feature extract sources, which weakens the effect of the classification.

Tab.4 Comparison of fault diagnosis of the six methods

Fault detection method and network structure	Accuracy of training set	Accuracy of testing set	Runnin g time/s
No feature extraction + DBN (13-100-20-10-2)	99.4 (6809/6850)	98(147/150)	21.4
No feature extraction + BP (13-50-2)	78.2 (5357/6850)	52.3(78/150)	43.2
No feature extraction + SVM (RBF, $c=16$, $g=0.9$)	81.4 (5576/6850)	66.7(100/150)	8.7
feature extraction + DBN (5-50-20-10-2)	99.8 (6836/6850)	96.6(145/150)	18.7
feature extraction + BP (5-50-2)	88.2(6042/6850)	74.63(112/150)	36.4
feature extraction + SVM (RBF, $c=16$, $g=0.9$)	89.4(6124/6850)	85.3(128/150)	6.4

RBF is Radial Basis Function, c is error costs , g is kernel function.

5. Summary

The paper built a big data-deep belief model for the fault detection of aero-engine air exhaust temperature sensors using the data adopted from the classification model based on DBN and flight parameter recorder. The fault detection model takes the 13-dimension data adopted from flight parameter recorder as input, and extracts features from high dimensional data automatically without manual intervention. The experiments prove that the features extracted by DBN are better than by manual intervention. It also can fully mine the data and avoid over-fitting during the process of deep network learning. There's no need of precise mathematical modeling and tedious artificial feature extraction when uses this model to detect the fault, so the model has an outstanding generalization effect that could be expended to the fault detection of other sensors and complex components. In spite of the advantages above, there is some problem like time-overspent, which will be mainly worked on in the future.

Acknowledgment

This research is partially supported by Aeronautical Science Foundation of China under Grant 20141396012.

Correspondence author:

Xiaofei Zheng:Authors Address is College of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an 710038, China;(e-mail:798981337@qq.com)

Author's brief introduction:

Chuang Guo:Received the Ph.D degree in College of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an.He is currently a associate professor with Air Force Engineering University.His current research interests include Vehicle health management and fault diagnosis.

Xiaofei Zheng:He is Currently pursuing the B.S degeree in College of Aeronautics and Astronautics Engineering, Air Force Engineering University, Xi'an. His current research interests include Vehicle health management and fault diagnosis.

References

- [1] WANG Xiuyan, LI Cuifang, GAO Mingyang, et al. Fault diagnosis of aero-engine gas based on SVM and SNN[J]. Journal of Aerospace Power, 2014,29(10):2493-2498. (in Chinese)
- [2] Prasanna Tamilselvan, Pingfeng Wang. Failure diagnosis using deep belief learning based health state classification [J]. Reliability Engineering and System Safety, 2013: 124-135.
- [3] HINTON G E, SALAKHUTDINOV R. Reducing the dimensionality of data with neural network [J]. Science, 2006, 313: 504-507.
- [4] BENGIO Y. Learning deep architectures for AI [J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.
- [5] AREL I,ROSE D C, KARNOWSKI T P. Deep machine learning: a new frontier in artificial intelligence research [J]. IEEE Computational Intelligence Magazine, 2010, 5(4): 13-18.
- [6] HINTON G E. Training products of experts by minimizing contrastive divergence [J]. Neural Computation, 2002, 14(8): 1771-1880.
- [7] GRAVES A, MOHAMED A R, HINTON G E. Speech recognition with deep recurrent neural networks[C]//Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference. Santa Clara:[s. n.]2013: 6645-6649.

- [8] XIE Chuan, NI Shihong, ZHANG Zongling. A New Method for Preprocessing in Absent Flight Parameter[J].Computer Simulation, 2005,22(4):27-31.(in Chinese)
- [9] LIU Jianwei, LIU Yuan, LUO Xionglin. Research and Development on Boltzmann Machine[J].Journal of Computer Research and Development,2014,51(1):1-16. (in Chinese)
- [10] HINTON G, OSINDERO S, THE Y.A fast learning algorithm for deep belief nets[J].Neural Computation,2006,18(7):1527-1554.
- [11] HINTON G, A Practical guide to training restricted Boltzmann machines [R]. Toronto: Machine Learning Group university of Toronto,2010;129-136.
- [12] HUANG Haibo, LI Renxian, YANG Qi, et al. Identifying Abnormal Noise of Vehicle Suspension Shock Absorber Based on Deep Belief Networks[J].JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY,2015,50 (5) : 776-782.