AMSE JOURNALS-2016-Series: Advances C; Vol. 71; N°1; pp 110-124
Submitted July 2016; Revised Oct. 15, 2016, Accepted Dec. 10, 2016

Wind Turbine Failure Risk Assessment Model Based on DBN

*Chen Fei, **Fu Zhongguang

* Energy & Power Engineering School, North China Electric Power University
China, Beijing, (chenfei@ncepu.edu.cn)
** Energy & Power Engineering School, North China Electric Power University
China, Beijing, (fzg@ncepu.edu.cn)

Abstract

As wind turbine is mainly composed of two strongly coordinated mechanisms, the transmission mechanism and the energy conversion, fault propagation characteristics and waveforms are fairly complex. Traditional Analysis Method of Kinetic Model, Expert System and Superficial Learning Model are effective in characteristic representation and failure analysis, but their prediction based on risk assessment is not adequately accurate. Furthermore, for big data which is multi-scale, heterogeneous, multi-source, modeling and training through those methologies is difficult. This paper proposes to apply DBN Depth Learning Theory to failure risk assessment of wind turbine. The experience from analysis on mechanical characteristics and image ones is deeply structured as the working machanism of human brain. Experiment indicates that the characteristic analysis method and failure risk assessment model based on DBN discussed in this paper has better performance in prediction accuracy and evolution ability than traditional solutions.

Key words

Wind Turbine, Failure Risk Assessment, Deep learning, Big data, DBN

1. Introduction

Wind turbines have higher failure rate than conventional generators, since they run outdoor in harsh natural environment over time and have to bear unpredictable reversal load and strong instant load sometimes [1]. Currently, most of the operation monitoring and maintenance systems in wind power farms are products introduced from other sectors. Though they are successful in other industries, they fall short of expectation in wind power industry [2-3].

In recent years, researchers have proposed a variety of feature extraction methods to solve classification problems. Features extracted include those in time domain, frequency domain, time-frequency and image. However, given the different natural environments (wind, magnetic, atmospheric pressure, humidity, rain, snow, etc.) and power grid conditions, operation of wind turbine creates big data which are heterogeneous, multi-type, and multi-scale [4]. Based on current theories and research methods, the mechanism, regular pattern and knowledge of turbine performance degradation was not mined and used at in-depth level on multiple scale[5][6]. This article, based on the Deep Learning Theory, discusses a method to identify the characteristics and judge the operation condition of wind turbines, so that manual feature extraction and selection process can be avoided, and the process is less complex and uncertain but smarter, leading to more accurate classification and prediction.

2. Theoretical basis

2.1 Analytical method of operating characteristic

(1). Mechanical characteristics

Wind turbines with periodic working parts as the main component, whose operating curve is affected by external environment. Those effects are mainly reflected in three dimensions, namely, time domain, frequency domain, and phases. Time domain shows working curve of mechanical components as time goes, frequency domain is the amplitude distribution in each frequency, phase means retardation of each waveform that makes up the curve, i.e. ratio of the time offset between waveform that takes up the period [7].

a) Fourier transform

Without external interference, the operation of wind turbine gear structure can be presented as the curve shown in figure 1. It can be stacked into different curve depending on the number of gears cooperating with each other.

Of course, this is one of the gear operation modes in the ideal scenario. Fourier transform can be used to decompose working curves into superimpose of several sine curves. After a sufficient number of learning samples, interference wave will be obtained, which is called de-noising.

$$F(w) = \mathrm{F}[f(t)] = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t}\, dt \qquad (1)$$

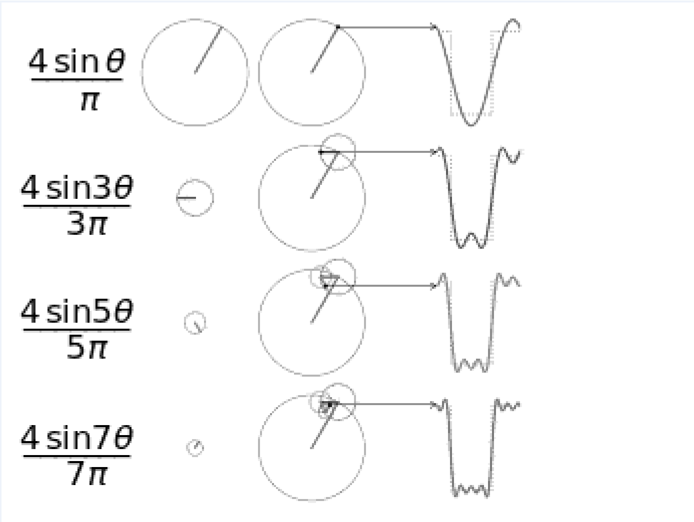

Fig.1. Working curve of the gear structure[8]

b) Wavelet transform

In grid environment, wind turbine is affected by external wind force and speed on the one hand, and internal structure such as capacitors and inductors on the other hand. Wind turbines present the fuzziness of periodic operation curve, and wavelet analysis of irregular change of waveform is more effective than a smooth sine wave. Wavelet transform can be understood as a substitute of sine wave in Fourier transform after wavelet function of scaling and translation. It is a localization mathematical transformation of a function in time domain and space domain. The time information of signal is obtained by translation of mother wavelet, and frequency characteristics of the signals is obtained through the scaling of mother wavelet. The law of wavelet superposition in the process of scaling and translating is found. Then the coefficients of wavelet of superposition, which expresses the local signal by coefficients and wavelet, is calculated [9][10].

Continuous wavelet transform (CWT) formula is as follows:

$$C(scale, position) = \int_{-\infty}^{+\infty} f(t)\varphi(scale, position, t)dt \tag{2}$$

(2). Image characteristics

a) Classic K-means clustering algorithm

Suppose the original data extracted is collection of (x_1, x_2, …, x_n), and each element xi is a vector with d dimensions, the purpose of K-means clustering is that, with the specific classification number k ($k \leq n$), the original data will be classified into k sets $S=\{S_1, S_2, \ldots, S_k\}$, expressed as follows inn numerical model, for the minimum [11]:

$$\sum_{i=1}^{k} \sum_{x_j \in S_i} ||x_j - u_i||^2 \quad (3)$$

Here u_i represents the average classification of S_i. As a cluster of pixels, S_i expresses identifiable point, line and surface. The image from wind turbine condition assessment usually reveals transmission structure corrosion, cracks, wheel peeling, missing teeth, displacement of coupler and bearings, blade crack etc.

The number of iterations of the K – means is, to a large extent, determined by the initial center and the number of clusters. The solution to the election issue of the center can be either Canopy algorithm or the Bayesian information criterion (BIC) algorithm. The former is faster but less accurate, while the latter is just the opposite [12][13].

In big data environment, Map-Reduce theory to Reduce K - means provide support to accelerate the convergence speed. The core idea is that: in map stage, original data will be mapped to different center. In reduction stage, samples belonging to the same center will be seen as a group, the average of a group will be the new center of the samples. Iteration will not stop until the center does not change.

b) PCA and SVM

In feature extraction model based on image recognition, calculation of the distance between the feature points involves multi-dimensional inner product computation. The complexity of K-means clustering algorithm has dimension disaster problem in multidimensional space. PCA and SVM is an algorithm where division of nonlinear characteristics in high-dimensional space can be mapped to the linear ones in low-dimensional space, so that feature learning can be in the high-dimension space, while computing cost is the same as that in low-dimension space, which makes high-dimensional feature extraction and learning possible [14][15].

Dimension transformation and reduction of dimension computation complexity depends on so called kernel function. Model and parameter of kernel function change the mapping of input and features extracted implicitly, which affects quality of features extracted, and ultimately changes the function and computational performance of kernel function. Usually, there are three kinds of generic kernel function [16].

Linear kernel function which is applied to linear separable problems:

$$k(x_1, x_2) = < x_1, x_2 > \quad (4)$$

Polynomial kernel function which is applied to map to (d+1) dimensional features space:

$$\mathrm{k}(x_1,x_2) = (< x_1, x_2 > +\mathrm{r})^d \tag{5}$$

Gaussian kernel function which is applied to mapping to any dimension feature space:

$$\mathrm{k}(x_1,x_2) = \exp(-\frac{\|x_1 - x_2\|^2}{2\delta^2})^2 \tag{6}$$

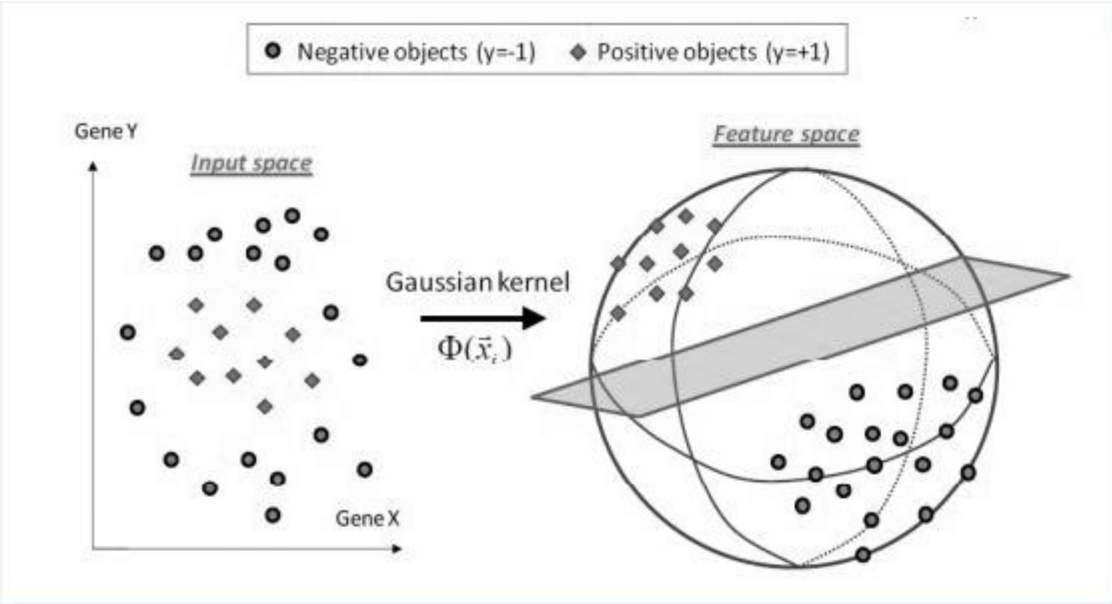

Fig.2. Gaussian kernel function transformation

2.2 RBM Algorithm model

The main idea of deep learning is to express knowledge hierarchically, and find essential characteristics of large number of data. At present, deep learning network model includes: Depth belief network model (DBN), Convolutional neural networks (CNN), Sparse Auto-Encoder (SAE), De-noising Auto-Encoder (DAE), etc. [17]. Among then, DBN is composed of several layer of Restricted Boltzmann Machine from bottom up, works through a greedy layer-by-layer learning algorithm, training process can be seen as the process of feature extraction. Statistical measures are extracted from all signals to represent the characteristics of the valve conditions, DBN extracts high-level knowledge based on the probability distribution theory, and others are based on primary features. Considering the complex structure of wind turbine, and primary features is fuzzy most of time, may be probability distribution theory have some superiority [18][19]. This paper chooses DBN based on RBM as deep learning model.

Restricted Boltzmann Machine (RBM) is a two-layer network model, as shown in figure 3, which is made up of the visible layer (v), hidden layer (h), and the connection matrix (w). Its core idea is that, visible layer v is used to represent a feature of the object being analyzed, the known v is converted to function p (v|h) to get new hidden layer h′, and p(h|v) transformation to get the new visual layer v′. By adjusting the parameter, the v′ is the same as the v, h′ is another expression of v, which can be used to represent the characteristic of the v[20].

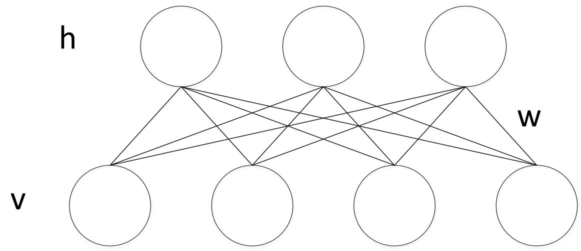

Fig.3. RBM model structure

$$p(h|v) = p(h_1|v)p(h_2|v)\ldots p(h_n|v) \tag{7}$$

$$p(v|h) = p(v_1|h)p(v_1|h)\ldots p(v_n|h) \tag{8}$$

a) RBM model is an undirected graph, assuming that the unit number of visible layer is n, unit number of hidden layer is m, and each pair of the visible and hidden units connected between the energy function is expressed as below:

$$E(\mathrm{v},\mathrm{h}|\theta) = -\sum_{i=1}^{n} a_i v_i - \sum_{j=1}^{m} b_j h_j - \sum_{i=1}^{n}\sum_{j=1}^{m} v_i w_{ij} h_j \tag{9}$$

Among them, vector v represents visible unit, vector h represents hidden unit, a_i and b_j represent the bias of element i in vector v and element j in vector h respectively, and w_{ij} represents the element in connection matrix of v and h.

b) For each pair of visible and hidden units in system model, joint probability distribution of energy function can be acquired in the case of energy function and the actual parameter $\theta=\{W_{ij},a_i,b_j\}$ to limit RBM is known.

$$P(\mathrm{v},\mathrm{h}|\theta)x = \frac{e^{-E(v,h|\theta)}}{Z(\theta)} \tag{10}$$

$$Z(\theta) = \sum_{\mathrm{v,h}} e^{-E(v,h|\theta)} \tag{11}$$

Of them, Z(θ) is called Partition Function, also namely the normalization factor.

c) The marginal distribution of joint probability distribution p (v,h|θ) can be obtained by:

$$p(v|\theta) = \frac{1}{Z(\theta)}\sum_{h} e^{-E(v,h|\theta)} \tag{12}$$

d) Since each unit in visible layer and hidden layer is independent from each other, the probability of all visible and hidden units can be acquired by the following formula:

$$P(v_i = 1|h) = \sigma(-\sum_i W_{ij} v_i - \sum_i b_i\, v_i) \quad (13)$$

$$P(h_i = 1|v\,) = \sigma(-\sum_j W_{ij}\, h_j - \sum_j a_j\, h_j) \quad (14)$$

σ(x) represents activation function, different kinds of activation function can be chosen according to the distribution of data in visible layer. For binary visible, the activation function can always be the following one.

$$\sigma(x) = \frac{1}{1 + exp(-x)} \quad (15)$$

3. Risk assessment model of wind turbine base on RBM

3.1 Characteristic analysis of risk assessment

Wind turbine generator mainly includes wind rotor system (including blades), drive system (including shaft, gear box and coupling), and yaw system (including the yaw bearing). Wind turbine blades could fail due to aging, screw loosening, cracks, and so on, and gear box and support bearing are likely to be imbalanced, misaligned, and rotor and stator rubbing could have faults such as loose, fixed structure [21][22].

a) Rotor imbalance

Physical characteristics: Harmonic energy of spectrogram focuses on the fundamental frequency, and unbalance vibration could trigger small frequency components. Below the critical rotor speed, the amplitude increases as the rotor speeds up; and the imbalance is relatively stable when the working speed is constant, at the same time, vibration is reflected in radial and axial, and axial vibration is greater than radial vibration. Phase remains the same, and synchronizes with the rotational speed.

Image characteristics: Rotor axis orbit is an ellipse, and the precession characteristics observed from axis orbit is synchronous precession.

b) Coupling misalignment

Physical characteristics: Coupling generates multiple frequency vibration, where parallel with misalignment produces twice frequency and angle misalignment is showed as highlights of the same frequency vibration. Coupling has the apparent phase characteristics, the phase difference in the same direction on both sides is 0° in case of parallel offset is misaligned, 180° in case of the angular misalignment, and between 0°~180° in case of integrated displacement.

Image characteristics: Due to the relative displacement of the axis and the axis of the coupling, the motion curve is cylinder in case of parallel displacement misalignment, and double-cone under the circumstance of angular displacement misalignment, and between in case of integrated displacement misalignment.

c) Dynamic and static friction of rotor system

Physical characteristics: When the waveform happens, the phenomena of single wave soldering "clipping", spectrogram could appear sum frequency and difference frequency between whirling frequency and rotational frequency. With the increase of contact arc, rotation frequency and amplitude decreases, and higher harmonics amplitude increases.

Image characteristics: The motion track of rotor is unstable, and changes to backward whirling from forward whirling.

d) Fixed structure loose

Physical characteristics: It has obvious phase characteristics with phase in line with rotational frequency, and the amplitude characteristic is that amplitude increases with the load size. Frequency content of the vibration signal is relatively complex, and it also produces higher harmonic and sub-harmonic vibration in addition to fundamental frequency.

Image characteristics: Orbit appeared deviation when not rotational structure loose happens, and different amplitude can be observed depends on the relative position between observation points and the structure when rotational structure loose happens.

3.2 Deep learning training model

a) Characteristic sampling model

Characteristic sampling model is based on Monte Carlo (MCMC) Gibbs sampling. The principle of MCMC shows that: through building the state transformation matrix, and iterating several times, the characteristics will be converged to steady state. The mathematical definition of Markov chain is as follows:

$$P(X_{t+1} = x \,|X_t, X_{t-1}, \ldots\ldots) = \mathrm{P}(X_{t+1} = x \,|X_t) \tag{16}$$

Gibbs sampling theory based on MCMC is as follows:

The probability distribution of bi-dimensional plane p(x, y) can be expressed as formula (17) and (18).

$$p(x_1,y_1)p(y_2|x_1)=p(x_1)p(y_1|x_1)p(y_2|x_1) \tag{17}$$

$$p(x_1,y_2)p(y_1|x_1)=p(x_1)p(y_2|x_1)p(y_1|x_1) \tag{18}$$

And they could be simplified to formula (19).

$$p(x_1,y_1)\,p(y_2|x_1)=p(x_1,y_2)p(y_1|x_1) \tag{19}$$

Which is:

$$p(\mathrm{A})p(y_2|x_1)=p(B)p(y_1|x_1) \tag{20}$$

In the similar way, take the same y of two points, the formula becomes (21).

$$p(\mathrm{A})p(y_2|x_1)=p(C)p(y_1|x_1) \tag{21}$$

Then the transition probability matrix between any two points in the plane could be constructed as formula (22)-(24).

$$\mathrm{Q}(\mathrm{A}\to\mathrm{B})=\mathrm{P}(y_B|x_1)\qquad condition{:}\,x_A=x_B=x_1 \tag{22}$$

$$\mathrm{Q}(\mathrm{A}\to\mathrm{C})=\mathrm{P}(x_C|y_1)\qquad condition{:}\,y_A=y_c=y_1 \tag{23}$$

$$\mathrm{Q}(\mathrm{A}\to\mathrm{D})=0\qquad condition{:}\,other \tag{24}$$

Any pair of points, X and Y in the plane can be derived according to Markov chain theory, which meets the carefully smooth conditions:

$$p(\mathrm{X})\mathrm{Q}(\mathrm{X}\to\mathrm{Y})=p(\mathrm{Y})\mathrm{Q}(\mathrm{Y}\to\mathrm{X}) \tag{25}$$

Of which, Markov chain in bi-dimensional plane is converged to stationary distribution p(x, y), and it could be extended to n-dimensional space, whose Gibbs sampling algorithm is shown as table1 [23].

Tab.1 Gibbs sampling algorithm

Algorithm 1：N-dimensional Gibbs sampling algorithm	
Step 1: Random initialization $\{x_i \mid i=1,2\ldots\mathrm{n}\}$	
Step 2: For t=0,1,2,3…n，sampling cycles:	
1.	$x_1^{t+1}\sim p(x_1\mid x_2^t,x_3^t,x_4^t\ldots\ldots\ldots\ldots\ldots\ldots\ldots x_n^t)$
2.	$x_2^{t+1}\sim p(x_2\mid x_1^{t+1},x_3^t,x_4^t\ldots\ldots\ldots\ldots\ldots\ldots x_n^t)$
	……
$j.$	$x_j^{t+1}\sim p\left(x_j\mid x_1^{t+1},x_2^{t+1},\ldots x_{j-1}^{t+1},x_{j+1}^t\ldots x_n^t\right)$
	……
$n-1.$	$x_{n-1}^{t+1}\sim p(x_{n-1}\mid x_1^{t+1},x_2^{t+1},\ldots\ldots\ldots x_{n-2}^{t+1},x_n^t)$
$n.$	$x_n^{t+1}\sim p(x_n\mid x_1^{t+1},x_2^{t+1},\ldots\ldots\ldots\ldots\ldots.x_{n-1}^{t+1})$

After a number of iterations, the probability distribution is converged as $p(x_1, x_2\ldots x_{n-1}, x_n)$, which is the steady-state sample.

b) Model parameters setting

As mentioned in section 2.2, the process of model parameters setting is to restrict solving actual argument RBM, $\theta=\{W_{ij},a_i,b_j\}$, and its likelihood assessment. In order to achieve optimal parameters, it needs to make the log-likelihood function $L(\theta)$ maximum, which is:

$$\theta' = \arg_{\theta} maxL(\theta) \tag{26}$$

The log-likelihood function $L(\theta)$ is defined as follows:

$$L(\theta) = \sum_{t=1}^{T} \log P(v^{(t)}|\theta) = \sum_{t=1}^{T} \log P(v^{(t)},h|\theta) = \sum_{t=1}^{T} \log \frac{\sum_h e^{-E(v^{(t)},h|\theta)}}{\sum_v \sum_h e^{-E(v,h|\theta)}} \tag{27}$$

Then the gradient of natural logarithm can be expressed as formula (28)

$$\frac{\partial L}{\partial \theta} = \sum_{t=1}^{T} \frac{\partial}{\partial \theta}\left[\log \frac{\sum_h e^{-E(v^{(t)},h|\theta)}}{\sum_v \sum_h e^{-E(v,h|\theta)}}\right] = \sum_{t=1}^{T}\left[\langle \frac{\partial(-E(v^{(t)},h|\theta))}{\partial \theta} \rangle_{P(h|v^t,\theta)} - \langle \frac{\partial(-E(v,h|\theta))}{\partial \theta} \rangle_{P(v|h,\theta)}\right] \tag{28}$$

Where $\langle\rangle_P$ is the average value of distribution P, and $p(h|v^t,\theta)$ is the probability distribution of implied layer. When the visible unit is qualified for learning sample v^t, first item of likelihood assessment function (formula 28) can be calculated, while the second item $p(v,h|\theta)$ is joint probability distribution and is hard to solve due to the normalization factor, it is needed to combine the MCMC sampling model, and calculate the approximate value of sample points.

Error value of likelihood between model and sample data is calculated by reconstruction error method, and the model most closely matches the sample data when the error is minimized.

Tab.2 Reconstruction error algorithm

Algorithm 2: Reconstruction error algorithm

Error = 0
For all $v^{(t)}$ do
    h-P(h|v)
    v-P(v|h)
    Error = Error + $\|\|v- v^{(t)}\|\|$
End for
Return Error

4 Application scenarios

The process of model training and risk assessment is shown as figure 4 (Gradient matching pursuit is applicable for pretreatment of image denoising[24]). By continuous iteration of feature data sampling and each RBM layer functions conversion, the process doesn't end until likelihood function meets the model assessment. If the assessment is inaccurate, it is needed to adjust the sample data selection and active RBM neurons of each layer, to achieve the goal of evolutionary neural networks.

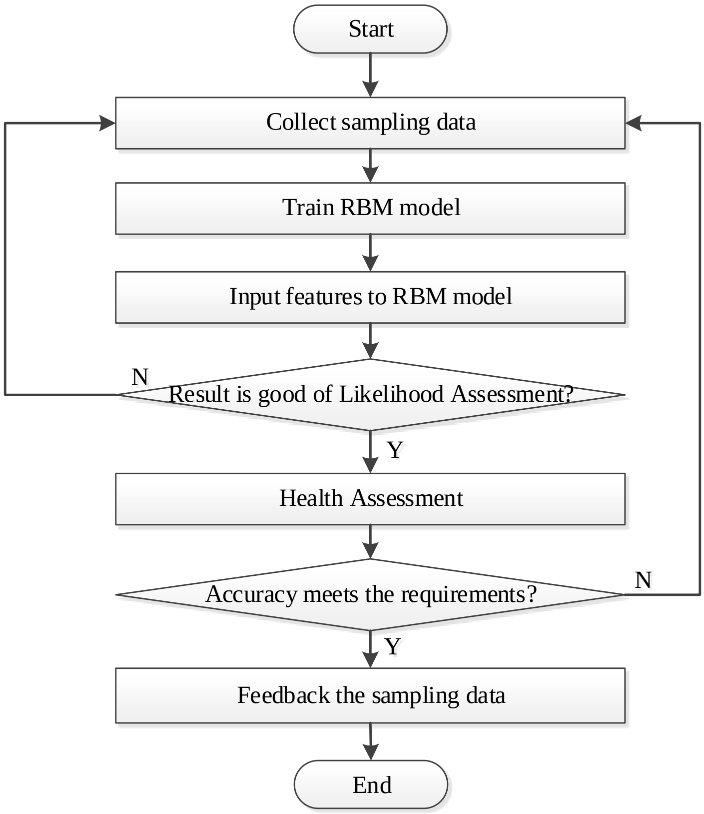

Fig.4. Deep Learning Process

Setting suitable RBM parameters could improve the precision of prediction and speed of model convergence effectively, and it is needed to study how to set the parameters through large iterative calculation, including layers of DBN model, hidden unit number of each layer, distribution of initial weights, punishment items of weights, times of training, and so on. The model in the iteration of parameters training should be assessed, to obtain the experience of parameters suitable for each scene, once the model is mature, and the real-time assessment could be certain.

This paper collects sampling data from a wind farm in a given region. The parameters includes 50 Visible Units expressed 50 important failure risk assessment indexes, 5000 Hidden Units of which every 100 characteristics can expressed one risk index. DBN was constructed by 20 RBM units, and the iteration times of DBN was 100, iteration times of model parameters training was 10000, learning rate was 0.1, initial weights was two decimal random number between 0 and 1, and the penalty factor was 0.01. Characteristic values were expressed in three-dimensional space, and the deep learning model was built with 20,000 iterations of Gibbs characteristics sampling. 100 wind turbines are taken as samples, and the probability of wind turbine risk status is analyzed to compare with the model prediction. The results were shown as figure 5, where dashes were predictive failure values, and solid lines were the failure rates of actual statistical analysis.

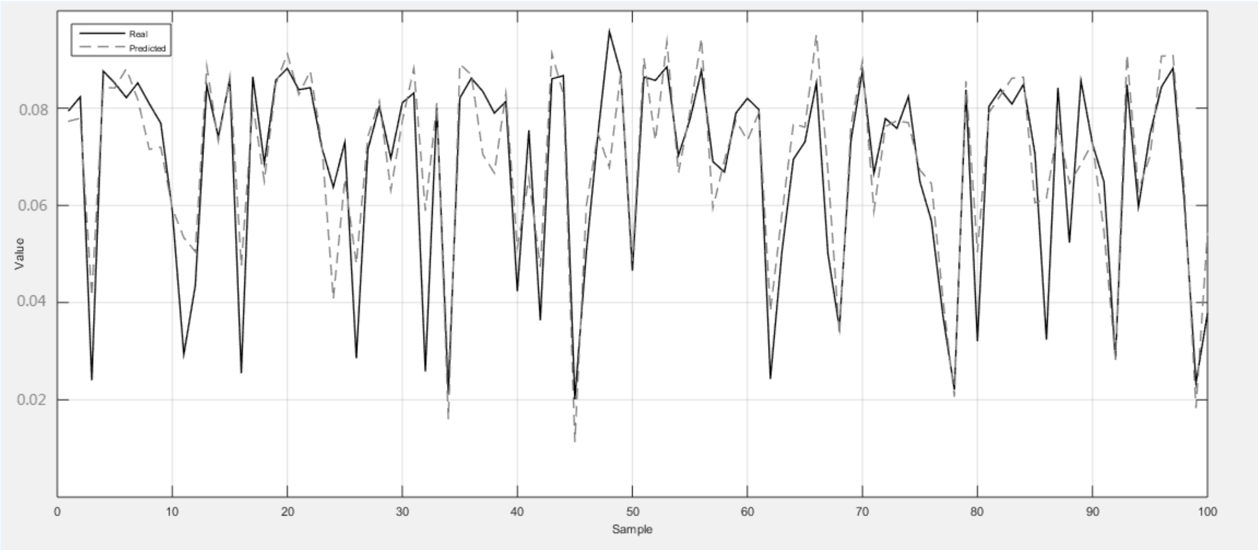

Fig.5. Comparison of precicted and actual risk

Conclusion

This paper firstly puts forward the method of wind turbine health assessment based on deep learning theory, combining with mechanical characteristics and image recognition, and designs the method of model validation and analysis. The application shows that failure risk assessment based on deep learning theory is better than traditional method of mechanical physical characteristic analysis and image clustering algorithm in several aspects, such as expression of complex and strong relevant problems, structure of experience, deepening and evolution of knowledge, and the accuracy of assessments.

Discussion

a) Why not use other deep learning model such as CNN, SAE,DAE?

Section 2.2 has answer this question, DBN extracts high-level knowledge based on the probability distribution of the sample, others are based on primary features. Considering the complex structure of wind turbine, and primary features is fuzzy most of time, may be probability distribution theory have some superiority.

b) What is the relationship between sampling and model parameter setting?

In section 3.2, log-likelihood function is mentioned to evaluate the quality of assessment model, but the function is hard to solve since there is a normalization factor, we use sampling data to obtain approximation. Quality of sample affects the result of log-likelihood function, ultimately affect the quality of the assessment model.

c) What is the innovation of this article?

This paper firstly put forward the method of wind turbine health assessment based on deep learning theory, combining with mechanical characteristics and image recognition, and designed the method of model validation and analysis.

d) How much accuracy has the solution discussed in this paper improved?

Model trained in this paper has an average error rate of lower than 5%, compared with the traditional method of about 90% accuracy, it’s a significantly improvement.

e) What is the direction of further research?

Parameter selection of training model is sightless, are there some methodology that is suitable for wind turbine failure risk assessment scenarios, it is a signification subject for further research.

References

1. He Dexin, Thinking of the Chinese wind power industry, 2008, High-tech and industrialization, vol.7, pp.25-27.
2. Crabtree C J, Feng Y, Tavner P J, Detecting incipient wind turbine gearbox failure: a signal analysis method for on-line condition monitoring, 2010, P Journal of Organic Chemistry, vol.75, no.18, pp.6122-6140.
3. Li Hui, Liu Shengquan, Ran Li, Li Yang,Hu Yaogang, Overview of Condition Monitoring Technologies of Power Converter for High Power Grid-Connected Wind Turbine Generator System, 2016, Transactions of China Electrotechnical Society, vol. 31, no. 8, pp.81-10.
4. Lei Yaguo, Jia Feng, Zhou Xin, Lin Jing. A Deep Learning-based Method for Machinery Health Monitoring with Big Data, 2015, Journal of Merchanical Engineering, vol. 51, no.21, pp. 49-56.
5. Zhu Qianwen, Ye Lin, Zhao Yongning, Lang Yansheng, Song Xuri,Methods for elimination and reconstruction of abnormal power data in wind farms, 2015, Power System Protection and Control , vol.3, pp. 38-45.

6. Jiang Cheng, Liu Wenxia, Zhang Jianhua, Yu Yue, Yu Jiaxi, Risk Assessment of Generation and Transmission Systems Considering Wind Power Penetration, 2014, Transactions of China Electrotechnical Society , vol.2, pp.260-270.
7. Heinrich. http://blog.jobbole.com/70549/. August 2016.
8. None.https://upload.wikimedia.org/wikipedia/commons/1/1a/Fourier_series_square_wave_circles_animation.gif. August 2016.
9. Meng Anbo, Chen Yucheng. Wind power combination forecasting based on wavelet packet transform and virtual forecasting method, 2014, Power System Protection and Control, vol.3, pp.71-76.
10. Qin Ling, Yang Jun. Applications of Wavelet Packet Analysis in the Fault Diagnosis of Rotating Machinery, 2014, Electronic Engineer, vol.27, no.2, pp.29-31.
11. Moondark.http://www.cnblogs.com/moondark/archive/2012/03/08/2385770.html.August 2016.
12. Zhao Qin, Efficient Algorithm of Canopy-Kmeans Based on Hadoop Platform. Electronic, 2014, Science and Technology, vol.27, no.2, pp.29-31.
13. Chu Yuezhong. An K-Means Clustering Algorithm Based on Bayesian Information Criterion, 2010, Journal of Anhui University of technology (natural science edition), vol.27, no.4, pp.409-412.
14. July. http://www.36dsj.com/archives/24596. August 2016.
15. Liao Jian, Shi Xianjun, Zhou Shaolei, Xiao Zhicai, Analog Circuit Fault Diagnosis Based on Local Graph Embedding Weighted-Penalty SVM, 2016, Transactions of China Electrotechnical Society, vol. 31, no.4, pp.28-35.
16. Liu Aiguo, Xue Yuntao, Hu Jianglu, Liu Luping. Ultra-short-term wind power forecasting based on SVM optimized by GA, 2015, Power System Protection and Control, vol.43, no.2, pp.90-95.
17. Luo Xianglong, Jiao Qinqin, Niu Liyao, Sun Zhuangwen Short-term traffic flow prediction based on deep learning, 2016, Application Research of Computers, vol.34.
18. Tong Shaomin. Research on Facial Activity Recognition based on Restricted Boltzmann Machine. Harbin, 2014, Harbin Institute of technology.
19. Van Tung Tran, Faisal AlThobiani, Andrew Ball. An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks,2014, Expert Systems with Applications, vol.41, pp.4113-4122.

20. Zhang Xin. Research on Music Classification Algorithm based on Deep Belief Network and Hidden Markov Mode, 2015, Dalian, Dalian university of technology.
21. Liu Xiaobo. Research and production of wind turbine fault monitoring and diagnosis system, 2009, Wuhan. Wuhan University of Science and Technology.
22. Ren Yan, Wu Qi-ren, Xue Li-ming. Health assessment of wind turbine, 2014, Advances in New and Renewable Energy, vol.2, no.6, pp.432~433.
23. Rickjin. http://cos.name/2013/01/lda-math-mcmc-and-gibbs-sampling/. August 2016.
24. Tang Ling, Chen Mingju. Image Denoising Method Using the Gradient Matching Pursuit, 2016, Mathematical Modelling of Engineering Problems. vol. 3, No. 2, pp. 53-56.