Advances in Computer, Signals and Systems (2020) 4: 1-6
Clausius Scientific Press, Canada

DOI: 10.23977/acss.2020.040101
ISSN 2371-8838

Fault diagnosis of wind turbine based on H-SKDB model

Baoyi Wang[1,a,*], Dongbing Yuan[1,b] and Shaomin Zhang[2,c]

[1]*School of Control and Computer Engineering,North China Electric Power University,Baoding 071003,China*
a. wangbaoyiqj@126.com, b.yuandb0103@163.com,
c. zhangshaomin@126.com

Keywords: Bayesian Network, Fault Diagnosis, Wind Turbine, H-SKDB

Abstract: The wind farm is in a bad wind area, which causes wind turbine faults to occur. Fault diagnosis of wind turbines is helpful for the maintenance and operation of wind turbines. The SKDB (Extensible Bayesian Network) model has the characteristics of high fault diagnosis accuracy and short training time. Based on the SKDB model, the Bayesian network structure is constructed by the method of mutual information addition calculation, then a fault diagnosis model of wind turbine based on H-SKDB is proposed, which realizes the fault diagnosis of wind turbine equipment information status. The experimental results show that the fault diagnosis method has higher calculation accuracy and shorter calculation time.

1. Introduction

Wind turbines are usually deployed in harsh wind farms, causing wind turbines to be prone to failure, which has a significant impact on the economics and safety of wind farms[1]. The fault of wind turbine can be found and maintained in time, which can effectively ensure the normal operation of wind farm.

In the field of wind turbine fault diagnosis, data-driven fault diagnosis methods have received more and more attention. Data driving refers to the use of power data in the SCADA system, combined with machine learning and artificial intelligence methods for data analysis, data classification, fault diagnosis and other operations. The SCADA system is integrated into the wind power plant, which stores a large amount of data related to the state of the wind turbine. A single wind turbine has hundreds of parameters such as voltage data, wind data, temperature data and time series data (such as vibration data) [3]. With the development of machine learning disciplines in recent years, ANN (Artificial Neural Network) has been applied in the field of wind turbine gearbox fault diagnosis [8], and ANN has achieved good fault diagnosis results with efficient processing capability. However, ANN still has great shortcomings because its results are too dependent on parameter adjustment and process optimization, and parameter adjustment and process optimization often require a lot of time and manual experience. In recent years, SVM (Support Vector Machine) has achieved good results in the field of gearbox, bearing and wind turbine fault diagnosis [9]. However, SVM can guarantee that the training time is within a reasonable range when processing small-scale data. When dealing with large-scale data, SVM will produce too long classification time and unstable classification effect.

Extended Bayesian Network (SKDB) is a class of Bayesian network classifiers. SKDB has the characteristics of fast convergence and short training time when dealing with massive data [2]. Compared with ANN-based wind turbine fault diagnosis methods, SKDB is more flexible and more interpretable, and can achieve better wind turbine fault diagnosis. Compared with SVM, SKDB has lower calculation time and training time, However, its processing accuracy still needs to be improved. In order to efficiently process the massive and heterogeneous power data collected by the SCADA system, based on SKDB and improve the network structure construction method, an H-SKDB fault diagnosis model is proposed, which has the characteristics of fault diagnosis accuracy and short training time. The Bayesian network constructed by H-SKDB can provide the detailed fault type probability information and the association information between each attribute feature variable, which has important reference significance for the design and implementation of wind turbine maintenance plans.

2. Problem Description

2.1.Data Source

The data used in this paper comes from the SCADA dataset and state dataset of a wind farm in China. The SCADA dataset is collected every 10 minutes, and the total number of samples is 74670, and the number of attribute features is 25. The state data set holds the corresponding wind power generation. Machine fault operation code (see Table 2 for detailed data). Table 1 only shows some parameters and data from March 17, 2018.

Table 1: Data collected by SCADA

time	Fan name	Wind speed	power	Generator speed	Generator front axle temperature	U_1 voltage	U_2 voltage	U_3 voltage
0:37:15	C1.241	12.45	1460	1754.9	8.09	414.0	414.0	414.3
0:47:32	C1.241	13.00	1452	1740.2	8.09	413.1	413.4	413.4
0:57:15	C1.241	13.31	1457	1757.8	8.12	412.8	413.1	413.1
1:05:02	C1.241	14.41	1433	1728.3	8.09	415.2	415.5	415.2
1:15:23	C1.241	14.26	1439	1754.9	8.08	415.5	416.1	416.1
1:26:30	C1.241	12.52	1413	1742.1	8.07	416.1	416.7	416.7

2.2.Wind Turbine Fault Type

Take the fault diagnosis of wind turbine as an example. The wind turbine operating status code is stored in the status data set and reflects the current operating state of the wind turbine. The generator running status code is recorded in the format of "main state: sub-state", a total of 175 types, and Table 2 lists some generator status codes. To simplify the discussion, 175 status codes are divided into six categories: (1) no fault; (2) power failure; (3) air-cooled fault; (4) feed fault; (5) generator over-temperature fault; 6) excitation fault, the above fault type will be used as a reference to wind turbine fault diagnosis results.

Table 2: Wind turbine status code

Date	Time	Status Code	Status Description
24/03/2018	12:37:38	0:0	Turbine in operation
14/05/2018	14:41:31	9:3	Generator heating: Hygrostat inverter
14/05/2018	19:49:23	2:1	Lack of wind: Wind speed to low

04/06/2018	08:17:53	60:11	Under-voltage L1 mains failure
05/06/2018	17:33:16	62: 505	No zero-crossing inverter
09/06/2018	00:00:00	228: 100	Timeout warn message
02/10/2018	23:42:04	80:21	Over-voltage DC-link

2.3.Fault diagnosis Process

Wind turbine fault diagnosis mainly includes: datapre-processing,network model construction, classifier fault identification, fault result output four parts, as shown in Figure 1. The classifier used in this paper is H-SKDB (Highly scalable k-order dependent Bayesian), the following describes the fault diagnosis process.

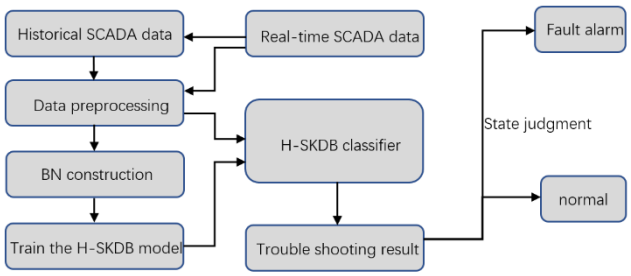

Figure 1: Fault diagnosis flowchart

1）Historical SCADA data and real-time SCADA data: SCADA contains a large number of multi-dimensional, heterogeneous wind turbine operating data, divided into historical data and real-time data usage. Historical data is used to train the construction of classifiers and Bayesian network; real-time data is pre-processed for online fault diagnosis.

2）Data pre-processing: In order to make a large number of multi-dimensional heterogeneous power data better used, the data is processed by normalization and under-sampling, and the data set is divided into training set and test set according to 7:3.

3）Bayesian network construction: H-SKDB belongs to BNC (Bayesian network classifier). For BNC of wind turbine fault diagnosis, the fault class label in historical SCADA data is used as the apex of the whole network structure. Then, the mutual information between each attribute variable Xi and the fault class Y and the mutual information between the respective Xi are calculated to generate a corresponding Bayesian network.

4）H-SKDB classification model: fault classification operation of real-time SCADA data, calculate the corresponding fault condition probability.

5）Fault diagnosis result: The diagnosis result is alarmed in various ways, such as voice playback alarm information and charts.

3. Fault Diagnosis Model Based On H-SKDB Algorithm

3.1 Data preprocessing

From the data distribution, the fault-free data accounts for 98.3% of the total number of samples, and the rest of the fault information accounts for about 1.7% of the total number of samples. For this data category imbalance problem, this paper divides the training set and test set into 7 :3, then use the under-sampling method to sample in the faultless data. Contains 921 fault data and 921 fault-free data,
394 fault data and 394 faultless data in the test set.

3.2 Fault Diagnosis Model

A total of 25 attributes of the data acquired in the SCADA system are denoted as X= {X_1, X_2...X_{25}}, and the corresponding data values are denoted as x= {x_1, x_2....x_n}, where n=25. According to the wind turbine state data set, the wind turbine fault state is divided into six categories, which are recorded as C= {C_1, C_2...C_6}, as shown in Table 3:

Table 3: Category 6 faults of wind turbines

Fault type	Fault free	Air cooling fault	Feed failure	Generator over temperature	Excitation fault	electricity failure
Numbering	C_1	C_2	C_3	C_4	C_5	C_6

The use of H-SKDB for fault diagnosis is mainly divided into two steps: 1) The network structure is constructed according to the mutual information addition calculation method.2) Calculate the condition probability P (C|x) according to the input data x, obtain the fault probability of the data sample, and obtain the fault classification result C.

When constructing the network structure, H-SKDB determines the dependence between variable nodes through mutual information. The mutual information is defined as follows: Let the joint distribution of two random variables (X, Y) be p (x, y) and the edge distribution be p(x), p(y), I (X; Y) is the relative entropy of the joint distribution p (x, y) and the edge distribution p(x)p(y), i.e.

$$I(X;Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{2-1}$$

For the attribute set X= {X_1, X_2, ...,X_n}, the mutual information between each variable X_i and the fault class variable C is sequentially calculated. Select the attribute of the top K of the mutual information value as the high correlation attribute random variable, select the maximum value of the mutual information as the root node X_{root},X_{root}=I (X_i|C) max, and then continue to calculate the mutual information to obtain Xi and X_j.The dependency between the mutual information is calculated as follows:

$$\sum_{j=1}^{q} I(X_i;X_j|C)\ P(c,x) \tag{2-2}$$

among them$Xj \in S, q = \min\{|S|;k\}$And add an arc between Xi and Xj in the network based on the relevant dependencies. After the network model is built, input the data x= {x_1, x_2, ...,x_k} and calculate the joint probability P (c,x):

$$P(c,x) = P(c)P(x_1|c)P(x_2|x_1,c)\dots P(x_n|x_1,\dots,x_{k-1},c) = P(c)\prod_{i=1}^{k} P(x_i|Pa_i,c) \tag{2-3}$$

Pa_i indicates that the Xi node in the network structure has a parent node other than the class variable, and P(c) is the fault prior probability. The conditional probability distribution of the node Xi in the network is expressed as P (x_i |Pa_i, c), which can be measured by the mutual information I (X_i; Pa_j), C). The mapping relationship between the two is:

$$\begin{aligned} P(x_i|Pa_i,c) &\Rightarrow I(X_i;Pa_i,C) \\ &= I(X_i;C) + i(X_i;Pa_i|C) \end{aligned} \tag{2-4}$$

After calculating the joint probability P (c, x), the conditional probability P (C|x) can be directly calculated:

$$P(C|x) = \frac{P(c,x)}{P(C)} \tag{2-5}$$

4 Experiment

4.1 Wind Turbine Over Temperature Fault Diagnosis Experimental Case

The experimental example of wind turbine over-temperature fault diagnosis is given below. Input wind turbine over-temperature fault data in training concentration, set over-temperature fault class variable to c_0. Bayesian network construction process is divided into two steps: 1) Calculate mutual information between all attributes and C_0, the attribute of the top five of the mutual information value is selected as the high correlation attribute set.X_0,$X_0 = \{X_9, X_{10}, X_{12}, X_{14}, X_{15}\}$, 2) Calculate the mutual information between the attributes and construct the network structure, as shown in Table 4.

Table 4: Mutual information between attributes

	Wind speed	Rotor speed	Generator front axle temperature	Generator rear axle temperature	Large generator temperature
Attribute number	9	10	12	14	15
9	7.529				
10	6.765	7.629			
12	6.720	6.821	7.588		
14	6.700	6.801	6.762	7.568	
15	7.114	7.216	7.175	7.154	7.982

According to the constructed wind turbine over-temperature network structure, input the maintenance period specified by the wind farmΔtThe internal training data is calculated according to the formula 2-3 to obtain the over temperature fault condition probability P(C|x), as shown in Figure 2.

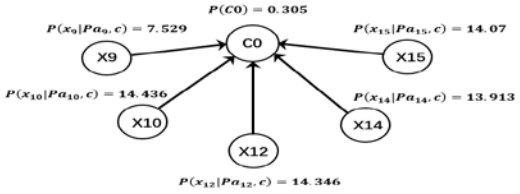

Figure 2: Generator over temperature fault condition probability

The above experiment calculates the over temperature fault condition probability P(C|x)=0.305, indicatingΔtThe probability of generator over-temperature fault occurring is 30.5%. The same method is adopted for the construction of other wind turbine fault network structures in the training set. The network structure and trainingdata are input into the H-SKDB fault diagnosis model, and finally The correct rate of fault diagnosis of wind turbines is shown in Table 5:

Table 5: Fault diagnosis Results

Fault number	Fault diagnosis correct rate%
C_0	93.2
C_1	94.1
C_2	94.7
C_3	96.1
C_4	97.6
average	95.1

4.2 Comparison of Fault Diagnosis Models

The training time and fault classification time can objectively reflect the performance of the fault diagnosis model. Figure 3 shows the training time and classification time of NB, SVM, ANN, SKDB, and H-SKDB under the same data set.

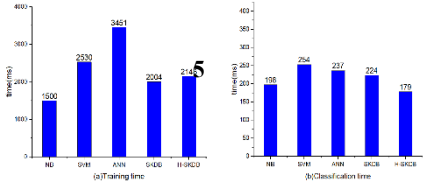

Figure 3: Training time and classification time comparison

Observed the results, the training time of H-SKDB is slightly higher than SKDB and NB, but the classification time is much lower than other classifiers. The fault diagnosis accuracy results are shown in Figure 4. Compare the classifications in Figure 3 and Figure 4. H-SKDB has lower training time and highest fault accuracy.

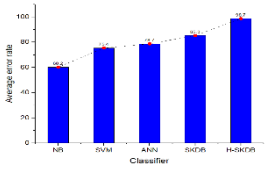

Figure 4: Comparison of fault diagnosis accuracy among various models

5 Conclusions

Aiming at the demand of wind turbine fault diagnosis under massive power data, this paper designs a fault diagnosis model based on H-SKDB. The wind turbine generator data from SCADA is used, and the SKDB model is improved. The Bayesian network structure of wind power data fault diagnosis based on mutual information addition calculation is constructed, and a new classifier H-SKDB is proposed. The experiment proves that the H-SKDB-based wind turbine fault diagnosis model can maintain a high fault diagnosis accuracy under the condition of ensuring low fault diagnosis time.

Reference

[1] *Y. Li, S. Liu, and L. Shu, "Wind turbine fault diagnosis based on Gaussian process classifiers applied to operational data," Renewable Energy, vol. 134, pp. 357–366, Apr. 2019.*

[2] *A. M. Martinez, G. I. Webb, S. Chen, and N. A. Zaidi, "Scalable Learning of Bayesian Network Classifiers," Journal of Machine Learning Research, vol. 17, no. 44, pp. 1–35, 2016.*

[3] *A. Lebranchu, S. Charbonnier, C. Bérenguer, and F. Prévost, "A combined mono- and multi-turbine approach for fault indicator synthesis and wind turbine monitoring using SCADA data," ISA Transactions, Dec.2019. 87: p. 272-281.*

[4] *T. S. Abdelgayed, W. G. Morsi, and T. S. Sidhu, "A New Approach for Fault Classification in Microgrids Using Optimal Wavelet Functions Matching Pursuit," IEEE Transactions on Smart Grid, vol. 9, no. 5, pp. 4838–4846, Sep. 2018.*

[5] *F. Petitjean, W. Buntine, G. I. Webb, and N. Zaidi, "Accurate parameter estimation for Bayesian network classifiers using hierarchical Dirichlet processes," Machine Learning, vol. 107, no. 8–10, pp. 1303–1331, Sep. 2018.*

[6] *J. Li, X. Zhang, X. Zhou, and L. Lu, "Reliability assessment of wind turbine bearing based on the Degradation-Hidden-Markov model," Renewable Energy, vol. 132, pp. 1076–1087, Mar. 2019.*

[7] *G. Helbing and M. Ritter, "Deep Learning for fault detection in wind turbines," Renewable and Sustainable Energy Reviews, vol. 98, pp. 189–198, Dec. 2018.*

[8] *J. Lei, C. Liu, and D. Jiang, "Fault diagnosis of wind turbine based on Long Short-term memory networks," Renewable Energy, vol. 133, pp. 422–432, Apr. 2019.*

[9] *S. Shi, B. Zhu, S. Mirsaeidi, and X. Dong, "Fault Classification for Transmission Lines Based on Group Sparse Representation," IEEE Transactions on Smart Grid, vol. 10, no. 4, pp. 4673-4682, July 2019.*

[10] *J.-Y. Zhou, F.-Y. Wang, and D.-J. Zeng, "Hierarchical Dirichlet Processes and Their Applications: A Survey: Hierarchical Dirichlet Processes and Their Applications: A Survey," Acta Automatica Sinica, vol. 37, no. 4, pp. 389–407, Jul. 2011.*

[11] *Zheng, Haiyang & Song, Zhe. (2009). Models for Monitoring of Wind Farm Power. Renewable Energy. 34. 583-590.*