

## Article

# A Failure Probability Calculation Method for Power Equipment Based on Multi-Characteristic Parameters

Hang Liu <sup>1,\*</sup>, Youyuan Wang <sup>1,\*</sup>, Yi Yang <sup>2</sup>, Ruijin Liao <sup>1</sup>, Yujie Geng <sup>2</sup> and Liwei Zhou <sup>1</sup>

<sup>1</sup> State Key Laboratory of Power Transmission Equipment & System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing 400044, China; rjliao@cqu.edu.cn (R.L.); willchou1202@cqu.edu.cn (L.Z.)

<sup>2</sup> Shandong Electric Power Research Institute, Shandong Electric Power Company, Jinan 250002, China; waiwai814@126.com (Y.Y.); gengyujie1225@126.com (Y.G.)

\* Correspondence: liuhangsheep@163.com (H.L.); y.wang@cqu.edu.cn (Y.W.); Tel.: +86-23-6511-1795 (Y.W.)

Academic Editor: Issouf Fofana

Received: 30 March 2017; Accepted: 11 May 2017; Published: 17 May 2017

**Abstract:** Although traditional fault diagnosis methods can qualitatively identify the failure modes for power equipment, it is difficult to evaluate the failure probability quantitatively. In this paper, a failure probability calculation method for power equipment based on multi-characteristic parameters is proposed. After collecting the historical data of different fault characteristic parameters, the distribution functions and the cumulative distribution functions of each parameter, which are applied to dispersing the parameters and calculating the differential warning values, are calculated by using the two-parameter Weibull model. To calculate the membership functions of parameters for each failure mode, the Apriori algorithm is chosen to mine the association rules between parameters and failure modes. After that, the failure probability of each failure mode is obtained by integrating the membership functions of different parameters by a weighted method, and the important weight of each parameter is calculated by the differential warning values. According to the failure probability calculation result, the series model is established to estimate the failure probability of the equipment. Finally, an application example for two 220 kV transformers is presented to show the detailed process of the method. Compared with traditional fault diagnosis methods, the calculation results not only identify the failure modes correctly, but also reflect the failure probability changing trend of the equipment accurately.

**Keywords:** failure probability; multi-characteristic parameters; the Weibull model; differential warning value; association rule; failure modes

## 1. Introduction

With the development of smart grid, the Chinese power industry is entering the era of Big Data [1–3]. Establishing the fault diagnosis models based on Big Data technology is of great importance since it can realize the comprehensive utilization of multi-source data and improve the accuracy of the fault diagnosis results, which is beneficial for ensuring reliability and safe operation, optimizing the condition-based maintenance (CBM) strategies, and increasing the remaining useful life (RUL) of power equipment.

With the development of online monitoring technology and information technology, the volume and variety of the condition monitoring data of the fault characteristic parameters have increased dramatically in recent years [4]. The fault diagnosis technology has become more automatic and intelligent than before by using multi-source data. For example, evidence fusion theory [5,6], support vector machine (SVM) [7], artificial neural network (ANN) [8–10], and other methods have been widely

applied to fault diagnosis for power equipment. Though many research achievements have been reached, there are still some shortcomings, which are summarized as follows:

- (1) Although the volume and variety of the monitoring data are large enough, most of the data changes in the normal range. That is to say, there are little fault data or abnormal data. However, many existing diagnosis methods often require a large amount of fault data to build and optimize the models. The low value density of the condition monitoring data limits the application of those methods [11–13].
- (2) Most existing methods have failed to reveal the association between the failure modes and various characteristic parameters. The equipment failure modes are difficult to identify correctly when some kinds of parameters are changed. Additionally, the values of some key parameters and factors in the models are difficult to estimate. Thus, they are always determined by empirical values or expert experience, which may cause large errors.
- (3) The insulation performance of power equipment is degraded gradually by the long-term impact of mechanical and electrical stress, which causes an increasing trend in the failure probability of the equipment. Many existing fault diagnosis methods can only qualitatively identify the failure modes of the equipment under abnormal conditions, but not quantitatively evaluate the failure probability and the changing trends of normal equipment. In fact, the equipment under normal conditions still has a risk of failure.

With the increasing reliability requirements of the power supply, the traditional fault diagnosis methods cannot meet the requirements of the fault diagnosis of the power system. How to make full use of the multi-characteristic parameters to improve the accuracy and the effectiveness when establishing the model becomes a significant challenge. This paper presents a failure probability calculation method for power equipment based on multi-characteristic parameters, which achieves the accurate identification of fault modes and quantifies the change trend of the failure probability of the power equipment.

This manuscript is organized as follows: In Section 2, the probability density functions and the cumulative distribution functions of different fault characteristic parameters are obtained, and the steps to disperse the parameters and calculate the differential warning values are presented. In Section 3, the method for association rules mining and the membership function calculations for parameters are given. In Section 4, the model to estimate the failure probability of the failure modes, as well as the equipment's failure probability, are introduced. In Section 5, the procedure of the method is listed in detail. An application example for two 220 kV power transformers is presented in Section 6, and the conclusions are drawn in Section 7.

## 2. The Discrete Intervals and the Differential Warning Values Calculation

### 2.1. The Weibull Model

The probability density function and the cumulative probability distribution function of the two-parameter Weibull distribution are shown in Equations (1) and (2), respectively [14].

$$f(x) = \frac{\beta}{\alpha} \left(\frac{x}{\alpha}\right)^{\beta-1} \exp \left[ -\left(\frac{x}{\alpha}\right)^{\beta} \right] \quad (1)$$

$$F = 1 - \exp \left[ -\left(\frac{x}{\alpha}\right)^{\beta} \right] \quad (2)$$

where  $x$  is the value of a kind of fault characteristic parameter;  $\alpha$  and  $\beta$  are the scale and shape parameters. When these two parameters are determined, the Weibull model is uniquely determined.  $\alpha$  and  $\beta$  can be estimated by using the maximum likelihood estimation method [15], and the steps are as follows:

- (1) The parameter time series  $X = [x_1, x_2, \dots, x_n]$  are Substituted into the logarithmic likelihood function equation shown in Equation (3). The likelihood functions are established in Equation (4):

$$\ln L(X) = \ln \prod_{i=1}^m \frac{\beta}{\alpha} \left( \frac{x_i}{\alpha} \right)^{\beta-1} \exp \left[ - \left( \frac{x_i}{\alpha} \right)^{\beta} \right] \quad (3)$$

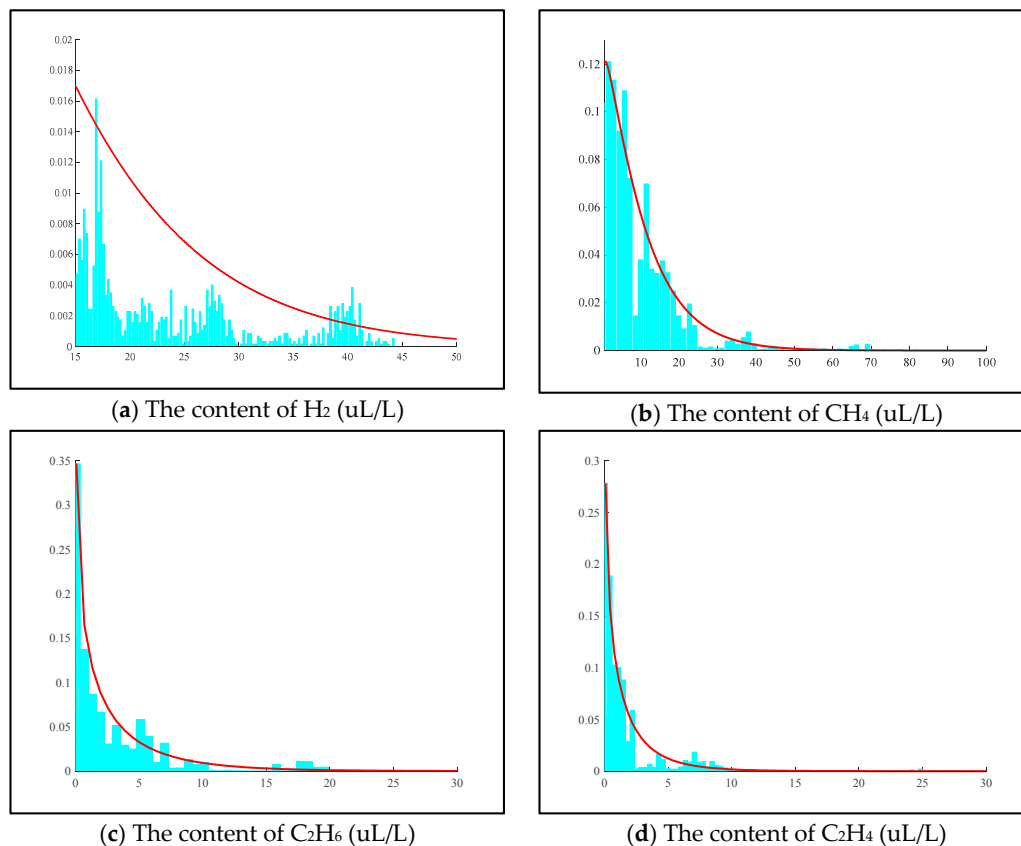
$$\begin{cases} \frac{\partial \ln(X)}{\partial \alpha} = 0 \\ \frac{\partial \ln(X)}{\partial \beta} = 0 \end{cases} \quad (4)$$

- (2) Substitute Equation (3) into Equation (4), then  $\alpha$  and  $\beta$  are calculated.

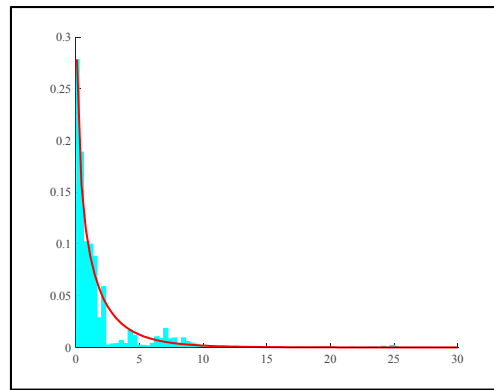
Taking the dissolved gas in transformer oil as an example, the dissolved gas content data in recent three years of different 220 kV transformers in a region were collected. The gas includes hydrogen, methane, ethane, ethylene, and acetylene, whose chemical formulae are  $H_2$ ,  $CH_4$ ,  $C_2H_6$ ,  $C_2H_4$ , and  $C_2H_2$ , respectively. According to the steps, the parameters of the Weibull model are estimated and shown in Table 1, and the probability density function curves and frequency histograms of the five types of gas are shown in Figure 1, respectively.

**Table 1.** The parameters calculation results of the Weibull model.

Parameters	$H_2$	$CH_4$	$C_2H_6$	$C_2H_4$	$C_2H_2$
$\alpha$	17.03	10.26	2.94	1.68	1.46
$\beta$	0.81	1.06	0.79	0.81	0.89



**Figure 1.** Cont.

(e) The content of C<sub>2</sub>H<sub>2</sub> (uL/L)**Figure 1.** The probability density function curves and frequency histograms of the dissolved gas.

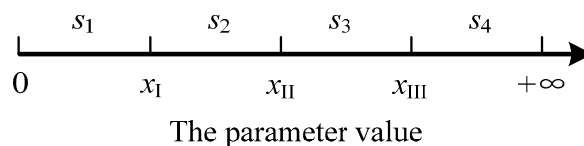
From Figure 1, the probability density functions are consistent with the actual distribution of each gas, which indicates the effectiveness of the Weibull model.

## 2.2. The Discrete Intervals Calculation

The fault characteristic parameters value is continuous, but the association rule mining algorithms require discrete input. Thus, the parameters should be discretized to different classes before data mining [16]. This paper presents a discrete method for parameters based on the distribution probability of the equipment condition. The steps are as follows:

- (1) According to [17], the equipment condition is divided into normal condition, attentive condition, abnormal condition, and serious condition. The distribution probability of the different conditions  $P = [p_{norm} \ p_{atten} \ p_{abn} \ p_{ser}]$  are calculated after collecting the actual condition data of all of the same type of equipment.
- (2) The cumulative distribution probability  $F = [F_I \ F_{II} \ F_{III} \ 1]$  is obtained by accumulating the distribution probability  $P$ .
- (3)  $F_I$ ,  $F_{II}$ , and  $F_{III}$  are brought into the inverse cumulative probability distribution function based on the two-parameter Weibull model expressed in Equation (5). Three values  $X_p = [x_I \ x_{II} \ x_{III}]$  are calculated and four discrete intervals  $s_1$ – $s_4$  are obtained, as shown in Figure 2.

$$x = \alpha[-\ln(1 - F)]^{1/\beta} \quad (5)$$

**Figure 2.** The discrete intervals of parameters.

The distribution probability and the cumulative probability of the 220 kV transformers are collected and listed in Table 2, and the calculation result of the discrete intervals of the five types of the gas are shown in Table 3.

**Table 2.** The data of 220kV transformers.

Equipment Condition	Normal Condition	Attentive Condition	Abnormal Condition	Serious Condition
Distribution probability	95.46%	3.53%	0.94%	0.07%
Cumulative probability	95.46%	98.99%	99.93%	1

**Table 3.** The discrete intervals of the gas (uL/L).

Gas	$x_I$	$x_{II}$	$x_{III}$
H <sub>2</sub>	82.44	139.89	257.84
CH <sub>4</sub>	29.75	43.22	66.57
C <sub>2</sub> H <sub>6</sub>	6.79	11.08	19.54
C <sub>2</sub> H <sub>4</sub>	5.19	8.10	13.55
C <sub>2</sub> H <sub>2</sub>	12.21	20.12	35.84

### 2.3. The Differential Warning Values Calculation

The abnormal increasing of fault characteristic parameters, such as the dissolved gas in oil, will increase the failure probability of the equipment. To ensure the safe and reliable operation of the power equipment, the warning value of each characteristic parameter is given by standard or by guide. If the parameter value exceeds the warning value, it indicates that the equipment is under poor condition or has a fault. For example, the warning value of hydrogen content is 150 uL/L, and the acetylene content is 5 uL/L for the 220 kV power transformers [17].

However, due to the operating environment, the service age and health status of each equipment are different, and the actual warning value of the equipment may deviate from the warning value in [17]. According to the method in [18], this paper presents a differential warning values calculation method for the equipment. The steps are as follows:

- (1) Count the average failure probability  $P_{ave}$  of the same types of equipment in a region.
- (2) The probability  $P_{ave}$  is substituted into the inverse cumulative probability distribution functions with different characteristic parameters, as shown in Equation (5), and the differential warning value  $V_{dif}$  of each parameter is obtained.

Table 4 is the differential warning value of the dissolved gas when the average failure probability of the 220 kV transformers is 0.18.

**Table 4.** The differential warning value of the gas (uL/L).

Gas	Differential Warning Value
H <sub>2</sub>	37.52
CH <sub>4</sub>	17.06
C <sub>2</sub> H <sub>6</sub>	3.27
C <sub>2</sub> H <sub>4</sub>	2.67
C <sub>2</sub> H <sub>2</sub>	5.80

## 3. Membership Function of the Fault Characteristic Parameter

### 3.1. The Association Rules Mining for Failure Modes and the Fault Characteristic Parameters

Different failure modes will have different effects on the change trend of the fault characteristic parameters, which are quantified by association rule mining algorithms, such as Apriori and FP-Growth algorithms [19,20]. The Apriori algorithm scans the whole database in each loop to calculate the support and confidence coefficient of the candidate item sets. The association rule mining by the Apriori algorithm includes the following steps.

- (1) The fault data, which have  $m$  kinds of fault characteristic parameters and  $g$  kinds of failure modes, are collected. All of the parameters are discretized in four classes.
- (2) The dataset of fault data is established and denoted as  $D_f = [G, d_{f1}, d_{f2}, \dots, d_{fm}]$ , where  $G$  is the transaction of the failure modes,  $G = [G(1), \dots, G(g)]$ . The  $d_{fi} = [d_{fi}(1), d_{fi}(2), d_{fi}(3), d_{fi}(4)]$  is the transaction of the fault characteristic parameter  $i$  ( $i = 1, \dots, m$ ).
- (3) The  $D_f$  is divided into  $m$  datasets according to each parameter, namely  $D_{f1} = [G, d_{f1}]$ ,  $D_{f2} = [G, d_{f2}]$ ,  $\dots$ ,  $D_{fm} = [G, d_{fm}]$ .
- (4) Set the minimum support ( $Supmin$ ) the minimum confidence coefficient ( $Confmin$ ), the Apriori algorithm is applied to finding out all of the frequent two-item sets  $G \rightarrow d_f$ , the support coefficient  $Sup(G \rightarrow d_f)$ , and the confidence coefficient  $Conf(G \rightarrow d_f)$  based on Equations (6) and (7):

$$Sup(G \rightarrow d_f) = \frac{\text{count}(G \rightarrow d_f)}{\text{count}(D_f)} \quad (6)$$

$$Conf(G \rightarrow d_f) = \frac{\text{count}(G \rightarrow d_f)}{\text{count}(D_f)} \quad (7)$$

where  $\text{Count}(\cdot)$  is the counting function of the item set.

Support is used to measure the statistical importance of association rules in the entire dataset. The greater the support, the more frequent the item sets or rules appear in  $D_f$ . If  $Sup(G \rightarrow d_f)$  is no less than the  $Supmin$ ,  $G \rightarrow d_f$  is regarded as a frequent item set. Otherwise,  $G \rightarrow d_f$  is regarded as an infrequent item set.

Confidence is used to measure the trustworthiness of association rules. If an association rule  $G \rightarrow d_f$  satisfies both  $Sup(G \rightarrow d_f) \geq Supmin$  and  $Conf(G \rightarrow d_f) \geq Confmin$ ,  $G \rightarrow d_f$  will be regarded as a strong association rule. Otherwise, it is a weak association rule.

### 3.2. Membership Function Calculation Method

According to association rules between the failure modes and the parameters, the distribution probability, that is, the confidence coefficient between each item set is obtained. For any kind of failure mode  $j$  and characteristic parameter  $i$ , Equation (8) is satisfied:

$$\sum_{k=1}^4 Conf(G(j) \rightarrow d_{fi}(k)) = 1 \quad (8)$$

The membership function of the characteristic parameter  $i$  for the failure mode  $j$  is expressed in Equation (9):

$$F_{ij} = \begin{cases} \frac{C_{ij}(1)}{x_I} \cdot x & x \in s_1 \\ C_{ij}(1) + \frac{C_{ij}(2)}{x_{II} - x_I} \cdot (x - x_I) & x \in s_2 \\ C_{ij}(1) + C_{ij}(2) + \frac{C_{ij}(3)}{x_{III} - x_{II}} \cdot (x - x_{II}) & x \in s_3 \\ C_{ij}(1) + C_{ij}(2) + C_{ij}(3) + \frac{C_{ij}(4)}{x_M - x_{III}} \cdot (x - x_{III}) & x \in s_4 \\ 1 & x > x_M \end{cases} \quad (9)$$

where  $C_{ij}(s_k) = conf(G(j) \rightarrow d_{fi}(s_k))$ ,  $x$  is the value of characteristic parameters.  $x_M$  is the value under the condition that the cumulative distribution probability is  $(F_{III} + 1)/2$ .

#### 4. The Failure Probability of Failure Modes and the Equipment

##### 4.1. The Failure Probability of the Failure Mode

According to the method referred to in Section 3, the membership function of different failure modes resulting from different fault characteristic parameters can be figured out. Those different membership functions have different contribution to the failure probability of the failure modes. Therefore, the failure probability of each failure mode is calculated by a weighted method. For example, the failure probability of the  $j$ th failure mode is calculated by Equation (10):

$$P_{G(j)} = F \cdot W = \sum_{i=1}^m F_{ij} \cdot \omega_{ij} \quad (10)$$

$$\sum_{i=1}^m \omega_{ij} = 1 \quad (11)$$

where  $F_{ij}$  and  $\omega_{ij}$  are the membership function and important weight from the  $i$ th characteristic parameter to the  $j$ th failure mode, respectively. The important weight  $\omega_{ij}$  of each fault characteristic parameter for different failure modes is calculated in Section 4.2.

##### 4.2. The Calculation for Important Weight

The existing weight calculation methods, such as analytic hierarchy process (AHP) [21], may lack accuracy because of the dependence of subjective judgments from experts. To ensure the rationality of the result, a method for weight calculation is presented according to the differential warning values of the characteristic parameters, which is shown as follows:

- (1) Obtain the differential warning values of the different characteristic parameters.
- (2) Count the number ( $N$ ) of samples in which the characteristic parameters are greater than, or equal to, their differential warning values shown in Equation (12):

$$N = \begin{bmatrix} N_{11} & N_{12} & \cdots & N_{1m} \\ N_{21} & N_{22} & \cdots & N_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ N_{m1} & N_{m2} & \cdots & N_{mg} \end{bmatrix} \quad (12)$$

For every  $N_{ij}$ , it is the number of samples of the  $i$ th characteristic parameter to the  $j$ th failure mode. The important weights of all of the fault characteristic parameters for the  $j$ th failure mode are obtained by the normalized calculation expressed in Equation (13):

$$\omega_{ij} = \frac{N_{ij}}{\sum_{i=1}^m N_{ij}} \quad (13)$$

According to Equations (12) and (13), the result of the importance weight  $W$  is obtained based on the calculation of the fault data and the differential warning value rather than the experience of experts so that the accuracy and rationality are ensured.

##### 4.3. The Failure Probability of the Equipment

Every type of the failure mode will cause equipment failure, and the series network is established between equipment failure probability and the failure probability of the failure modes [14]. The calculation method of the equipment failure probability is given by Equation (14):

$$P_{equ} = 1 - \sum_{j=1}^g (1 - P_j) \quad (14)$$

where  $P_{equ}$  is the equipment failure probability, and  $P_j$  is the failure probability of the  $j$ th failure mode.

## 5. The Procedure of the Method

The flow of the method is shown in Figure 3. The key steps are as follows:

- (1) According to the method in Section 2, the probability density functions and cumulative distribution functions of different fault characteristic parameters based on a two-parameter Weibull distribution model are calculated. Based on the inverse cumulative distribution function, the condition distribution probability and the average failure probability of the equipment are applied to calculating the discrete intervals and the differential warning value of each characteristic parameter.
- (2) According to the method in Section 3, the Apriori algorithm is applied to finding the association rules in the fault data between failure modes, as well as different parameters and calculating of the confidence coefficient of the frequent item sets. After that, the cumulative probability distribution function of each characteristic parameter is obtained.
- (3) The membership function and the important weight of each parameter are obtained in Section 4.
- (4) The failure probability of all of the failure modes are calculated by a weighted calculation, and the series model is proposed to calculate the failure probability of the equipment.

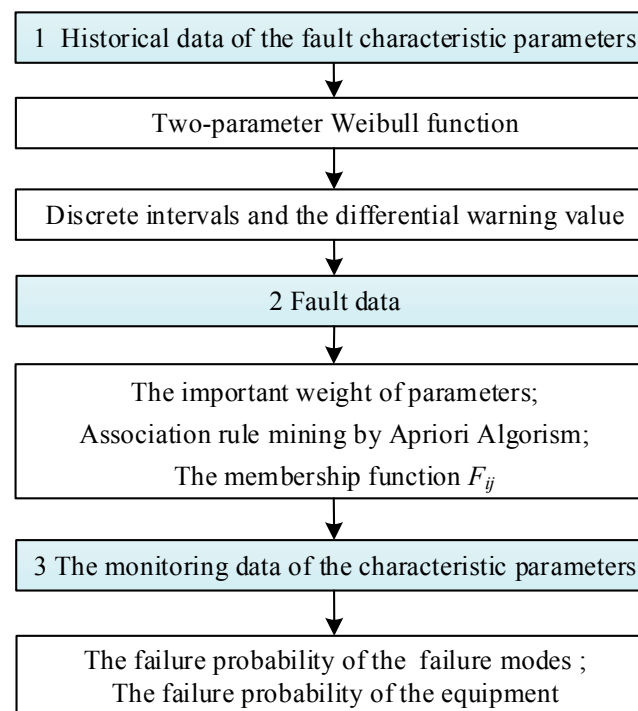


Figure 3. The flow of the failure probability calculation.

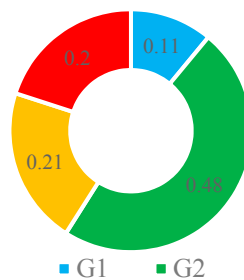
## 6. Application Example

### 6.1. Basic Information

According to the fault data samples of the 220 kV transformers, four types of the failure modes are defined, namely, low-temperature overheating ( $G_1$ ), high-temperature overheating ( $G_2$ ), low-energy



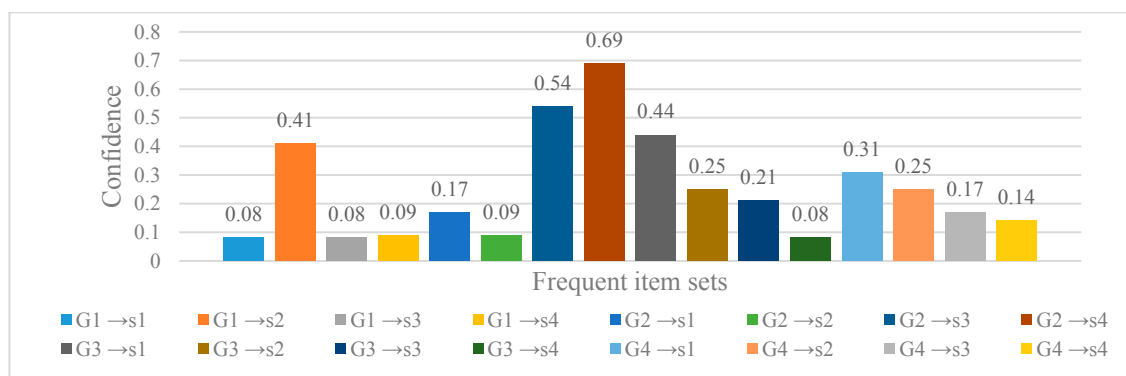
discharging ( $G_3$ ), and high-energy discharging ( $G_4$ ). The quantitative distribution of each failure mode is shown in Figure 4.



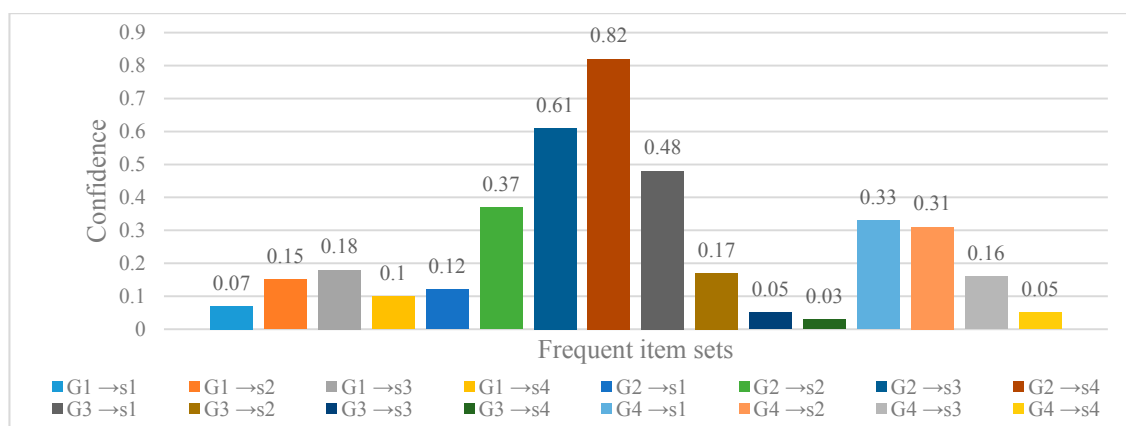
**Figure 4.** The quantitative distribution of each failure mode.

## 6.2. The Failure Probability and Important Weight

After setting the minimum support  $Sup_{min} = 0.02$  and the minimum confidence  $Conf_{min} = 0.02$ , the association rules among all of the failure modes and the different ways by which the gas dissolves are calculated by Apriori algorithm. Taking methane and ethane as examples, the frequent item sets and the confidence results are shown in Figure 5a,b.



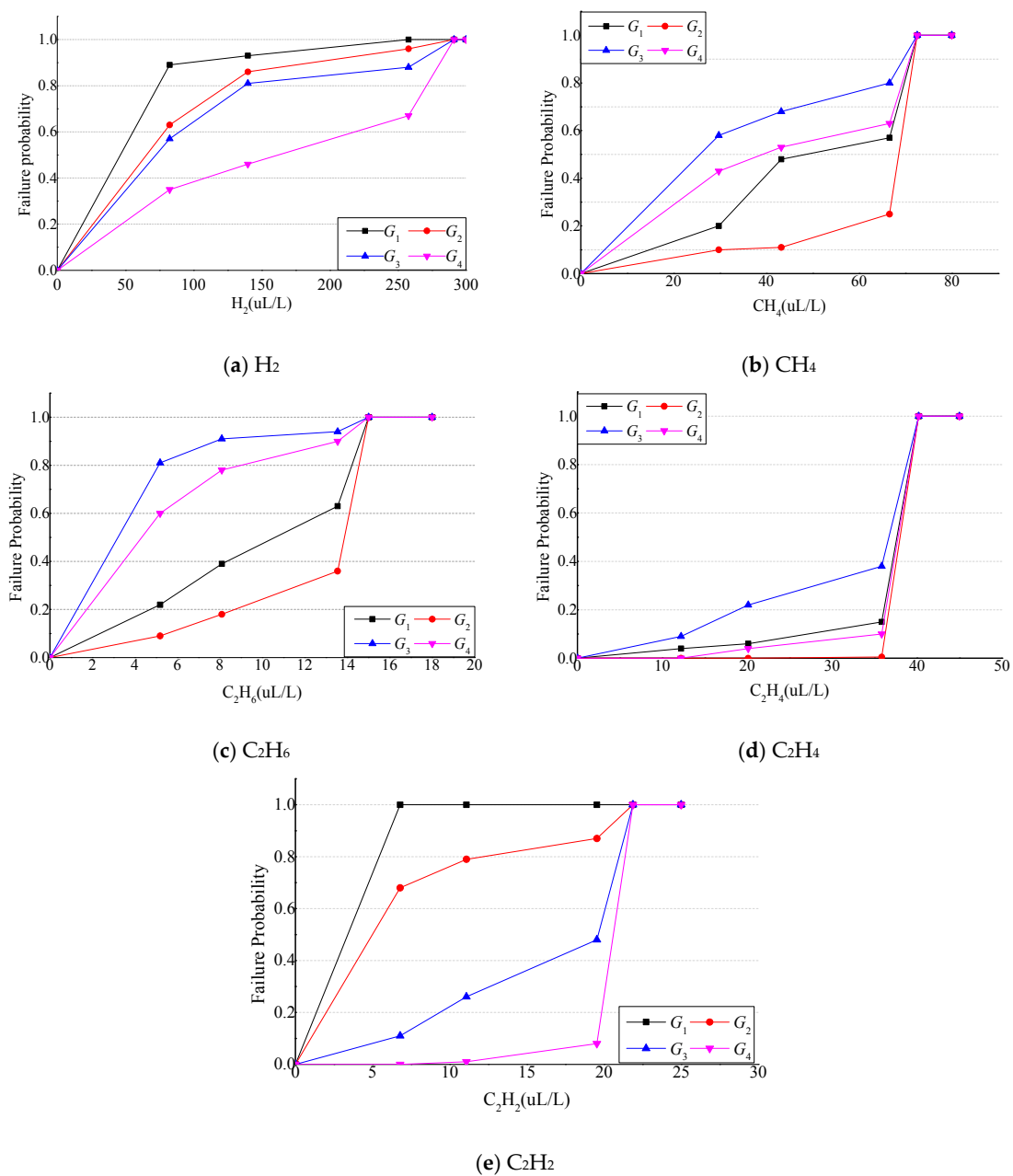
**(a)** The association rule between failure modes and methane.



**(b)** The association rule between failure modes and ethane.

**Figure 5.** The calculation result.

The membership functions of different kinds of gas are calculated, and the curves of the functions are shown in Figure 6.



**Figure 6.** The membership functions of the dissolved gas.

The important weights of the five types of gas to the four failure modes are calculated and shown in Figure 7.

It is obvious that the weight of acetylene is small in overheating failure modes while the alkanes is larger. Additionally, the weight of hydrogen is much larger than the weight of other gases in discharging failure modes.

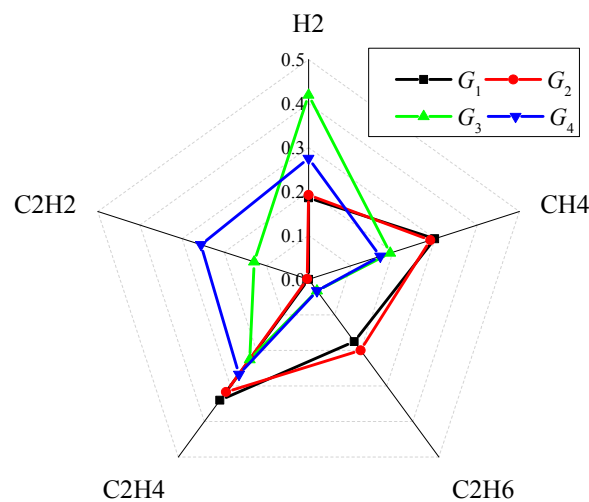


Figure 7. The important weight of each gas.

### 6.3. Case Study 1

Shown in Figure 8 are the time series data, which include 236 samples from 1 January 2016 to 30 June 2016, of the dissolved gas content of a 220 kV transformer (named T1) where the value of ethylene is zero.

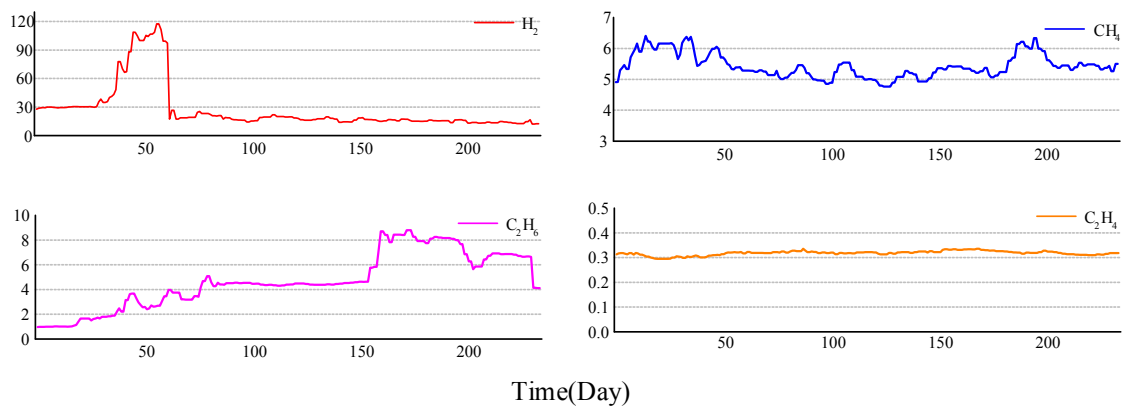


Figure 8. The content of gas (uL/L).

The data of the five kinds of dissolved gas are brought to each kind of membership function, and the failure probability of the four failure modes and the failure probability of T1 are calculated by Equations (10) and (14). The calculation result is shown in Figure 9.

It is clear from Figure 9 that the hydrogen content in T1 abnormally increased from the 45th day to the 63rd day, and surpassed the warning value (37.52 uL/L) so that the failure probability of the low-energy discharging ( $G_3$ ) and low-temperature overheating ( $G_1$ ) increased, resulting in the increment of the equipment failure probability. With the hydrogen content decreasing after the 63rd day, the equipment failure probability decreased to a lower level, and remained lower than 0.3. The calculation results are consistent with the actual situation of T1.

### 6.4. Case Study 2

The monitoring data of a 220 kV transformer (named T2) in 2015 are listed in Table 5. The overheat defect was detected in the oil pump on the 96th day, and it became serious after the 99th day. The diagnosis resulting from three-ratio method is 022 which means the failure mode of T2 is

high-temperature overheating ( $G_2$ ). The content of total hydrocarbon is the sum of  $\text{CH}_4$ ,  $\text{C}_2\text{H}_6$ ,  $\text{C}_2\text{H}_4$  and  $\text{C}_2\text{H}_2$ .

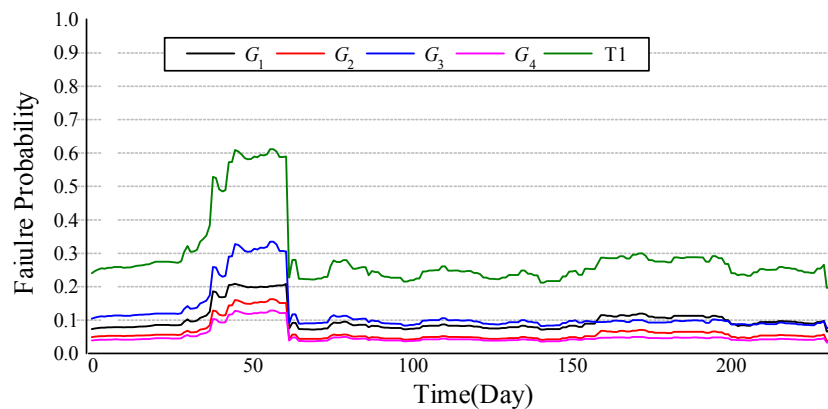


Figure 9. The failure probability calculation result of T1.

Table 5. The content of the dissolved gas in T2 (uL/L).

Day	$\text{H}_2$	$\text{CH}_4$	$\text{C}_2\text{H}_6$	$\text{C}_2\text{H}_4$	$\text{C}_2\text{H}_2$	Total Hydrocarbon
1st	0	1.6	0	0.6	0	2.2
31st	14	6.6	26	1.4	0	34
96th	73	110	41	190	0	341
99th	77	120	44	210	0.8	374.8
105th	95	130	55	220	1.2	406.2
107th	76	140	64	230	1.1	435.1
110th	86	137	48	230	1.1	416.1
146th	135	230	83	320	1.4	634.4
196st	140	240	85	410	1.1	736.1

The failure probability calculation result of T2 is shown in Figure 10, which indicates that failure probability was increasing rapidly from the 31th day to the 96th day, and the failure mode of T2 was  $G_1$  or  $G_2$ . After the 96th day, the failure probability of T2 approached 1, which means that T2 had a serious overheating fault. The result is consistent with the actual situation.

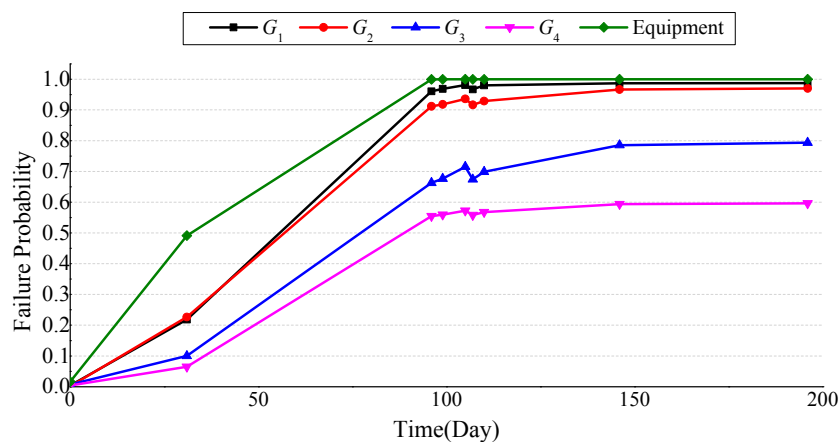


Figure 10. The result of failure probability calculation.

### 6.5. Comparative Analysis

To make a comparison between the method in this paper and the traditional methods, this paper presents two types of common fault diagnosis methods based on classifier and the threshold values.

#### 6.5.1. The Fault Diagnosis Method Based on a Classifier

Three kinds of common classifiers are presented to establish the fault diagnosis models, respectively, namely, decision tree classifier (J48), random forest classifier (RF), and SVM. After the training process of each model, the accuracy results of the three models are obtained and listed in Table 6. According to the precision results, the models based on RF and SVM are selected for fault diagnosis.

**Table 6.** Precision results of each model.

Day	Precision of the Model
J48	0.873
RF	0.997
SVM	0.952

#### 6.5.2. The Fault Diagnosis Method Based on the Threshold Value

The content and the relative increase rate of the dissolved gas in oil are important criteria for fault diagnosis. The threshold values of different kinds of gas are listed in Table 7 according to [17].

**Table 7.** The threshold values of the dissolved gas in oil.

Fault Severity	The Fault Criterion
Slight fault	The content of total hydrocarbon or $H_2$ is greater than 150 $\mu\text{L/L}$ .
Serious fault	The content of $C_2H_2$ is greater than 5 $\mu\text{L/L}$ ;
	The relative increase rate of total hydrocarbon is greater than 10% per month; The increase rate of total hydrocarbon is greater than 10 mL per day.

The monitoring data of Case 1, shown in Figure 8, is applied to the fault diagnosis model based on RF, SVM, and the threshold value. The results indicate the equipment is in a normal state without any fault, which is consistent with the actual situation. Similarly, the fault diagnosis result of Case 2 is also obtained and listed in Table 8. It indicates that the failure mode of the T2 is the high-temperature overheating fault ( $G_2$ ), and it becomes serious after the 96th day.

**Table 8.** The fault diagnosis results of three models.

Day	RF	SVM	Guide
1st	Normal	Normal	Normal
31st	Normal	Normal	Normal
96th~196st	$G_2$	$G_2$	Serious

According to the results, three traditional fault diagnosis methods have high accuracy in determining whether the equipment fails. However, the model of the threshold value fails to identify the failure modes for the equipment while the method based on the classifier is unable to determine the severity of the fault. Furthermore, any traditional methods can not quantify the failure probability of an equipment failure and reflect the changing trend of the failure probability. Taking T2 as an example, the fault cannot be found before the 96th day by traditional methods, so that maintenance strategy for T2 is not be made or arranged in time. The equipment failure may cause serious power interruption,

and significant losses of equipment and to the power system. These shortcomings of the traditional models make it impossible to meet the effectiveness requirements for the power system.

## 7. Conclusions

In this paper, a failure probability calculation method for power equipment based on multi-characteristic parameters is proposed. The historical data of different fault characteristic parameters are collected to obtain the cumulative distribution function and the differential warning value of each parameter. Then the association relations between different failure modes and the characteristic parameter and the important weight of each parameter are obtained. Finally, the weighted calculation and the series model are applied to calculating the failure probability of all of the failure modes and the equipment.

The failure probability calculation method realizes the comprehensive utilization of multi-characteristic parameters. To avoid the influence of subjective factors on the calculation results, the process of the fault probability function calculation is driven by data, but not judgment from experts, so that the results are ensured.

An application example for two 220 kV transformers is presented, and the procedure of the method is described in detail. The results indicate that this novel method has obvious advantages in accurately quantifying the failure probability, reflecting the failure changing trend of the equipment and helping to create a timely maintenance strategy so that the remaining useful life and the reliability of the power equipment can be markedly increased.

**Acknowledgments:** The work is supported by the National High-tech R and D Program of China (863 Program, 2015AA050204) and the China State Grid Corporation of Science and Technology Project (520626150032).

**Author Contributions:** The research presented in this paper was a collaborative effort among all authors. Hang Liu, Youyuan Wang, and Liwei Zhou proposed the methodology of the model and contributed to the manuscript at all stages. Ruijin Liao discussed the results and revised the manuscript critically. Yi Yang and Yujie Geng prepared the data for the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Chinese Society for Electrical Engineering Information Committee. *Chinese Electric Power Big Data Development White Paper*; China Electric Power Press: Beijing, China, 2013; pp. 1–13.
2. Yang, Y.D.; Bi, Z.Q. Advances and future challenges in electric power big data. In Proceedings of the 2014 Second International Conference on Advanced Cloud and Big Data, Wuhan, China, 20–22 November 2014.
3. Zhang, P.; Yang, H.F.; Xu, Y.B. Power big data and its application scenarios in power grid. *Proc. CSEE* **2014**, *34*, 85–92. [[CrossRef](#)]
4. Lai, Z.T. *Electric Power Big Data: Energy Internet Age Transformation and Value Creation*; Machinery Industry Press: Beijing, China, 2015; pp. 123–131. ISBN: 978-7-111-51693-4.
5. Otman, B.; Yuan, X.H. Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory. *Inf. Fusion* **2007**, *8*, 379–386. [[CrossRef](#)]
6. Lu, F.; Jiang, C.Y.; Huang, J.Q. A novel data hierarchical fusion method for gas turbine engine performance fault diagnosis. *Energies* **2016**, *9*, 828. [[CrossRef](#)]
7. Li, J.; Zhang, Q.; Wang, K. Optimal dissolved gas ratios selected by genetic algorithm for power transformer fault diagnosis based on support vector machine. *IEEE Trans. Dielectri. Electri. Insul.* **2016**, *23*, 1198–1206. [[CrossRef](#)]
8. Lu, J.J.; Huang, J.Q.; Lu, F. Sensor fault diagnosis for aero engine based on online sequential extreme learning machine with memory principle. *Energies* **2017**, *10*, 39. [[CrossRef](#)]
9. Hazlee, A.I.; Xin, R.C.; Ab, H.A.B. Hybrid modified evolutionary particle swarm optimisation-time varying acceleration coefficient-artificial neural network for power transformer fault diagnosis. *Measurement* **2016**, *90*, 94–102. [[CrossRef](#)]
10. Chen, W.G.; Pan, C.; Yun, Y.X.; Liu, Y.L. Wavelet networks in power transformers diagnosis using dissolved gas analysis. *IEEE Trans. Power Deliv.* **2009**, *24*, 187–194. [[CrossRef](#)]

11. The “four Vs” of Big Data. Available online: [http://www.ibmbigdatahub.com/sites/default/files/infographic\\_file/4-Vs-of-big-data](http://www.ibmbigdatahub.com/sites/default/files/infographic_file/4-Vs-of-big-data) (accessed on 12 October 2012).
12. Peng, X.S.; Deng, D.Y.; Cheng, S.J.; Wen, J.Y.; Li, C.H.; Niu, L. Key technologies of electric power big data and its application prospects in smart grid. *Proc. CSEE* **2015**, *35*, 503–511. [[CrossRef](#)]
13. Song, Y.Q.; Zhou, G.L.; Zhu, Y.L. Present status and challenges of big data processing in smart grid. *Power Syst. Technol.* **2013**, *37*, 928–935. [[CrossRef](#)]
14. Li, W.Y. A.3.4 Weibull Distribution. In *Risk Assessment of Power Systems: Models, Methods, and Applications*; John Wiley & Sons Press: New York, NY, USA, 2005; ISBN: 1118843355.
15. Zhong, B.; Liu, Q.S.; Liu, C.L. *Mathematical Statistics*; Higher Education Press: Beijing, China, 2015; pp. 103–115.
16. Zheng, Y.B. Research on Ensemble Learning of Fault Diagnosis and Prediction and Maintenance Decision-Making Models for Transformers. Ph.D. Thesis, Chongqing University, Chongqing, China, October 2011.
17. The State Grid Corporation. *Guide for Evaluation of Oil-immersed Power Transformers (Reactors)*; The State Grid Corporation: Beijing, China, 2008.
18. Qi, B.; Zhang, P. Calculation Method on differentiated warning value of power transformer based on distribution model. *High Volt. Eng.* **2017**, *42*, 2290–2298. [[CrossRef](#)]
19. Sheng, G.H.; Hou, H.J.; Jiang, X.C. A novel association rule mining method of big data for power transformers state parameters based on probabilistic graph model. *IEEE Trans. Smart Grid* **2016**, *1*, 1–9. [[CrossRef](#)]
20. Suresh, J.; Rushyanth, P.; Trinath, C. Generating associations rule mining using Apriori and FPGrowth Algorithms. *Int. J. Comput. Trends Technol.* **2013**, *4*, 887–892.
21. Xu, Z.H.; Wei, C.P. A consistency improving method in analytic hierarchy process. *Eur. J. Oper. Res.* **1999**, *116*, 443–449. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).