

# Transformer Defect Data Analysis Based on Data Mining Technology

Feng Li<sup>1</sup>, Hongbin Wang<sup>1</sup> and Kun Feng<sup>2</sup>

<sup>1</sup>Electric Power Research Institute of Guangdong Power Grid Corporation, GuangZhou, GuangDong, 510080, China

<sup>2</sup>China Real-time Technology Corporation, Beijing100192, China

**Abstract**—Using data mining technology to transformer operation analysis of the defects in the process of data mining, mainly for the entire network analysis of the defect data transformer account, defects, fault, environment and other information. Analysis mainly from the state information of equipment, equipment status, the relationship with their own attribute and the external environment, in order to obtain the associated factors influencing the equipment status, for the maintenance strategy optimization, equipment update strategy, manufacturers technical evaluation and other production and management decision-making to provide information support.

**Keywords**-data mining; transformer; defect data

## I. INTRODUCTION

The main transformer substation is one of the important equipment in power grid. Catastrophic failure of the main transformer often results in serious consequences and causes an emergency power outage events. The continuing power outage not only affects the people's daily life but also harms the quality of life. Therefore, during the main transformer selection, according to the different regions and climate, how to select the transformer is important, especially for the safety of the power grid. By analyzing a large number of main transformer operation defect data mining, this passage finds the fault conditions of various types of transformers obtained under various regions and in different climates, and provides reference and guidance for the selection and maintenance of transformer in different regions and climates.

## II. DATA MINING TECHNOLOGY BASED ON ASSOCIATION RULES

Data mining is analyzing data source by some way, where you can find some potentially useful information, so it also called data mining and knowledge discovery. And association rule mining is a very important issue in data mining, as the name suggests, it is found a correlation which may exist between things or contact from behind the data.

The basic principles of the association rules: Let  $I=\{i_1, i_2, \dots, i_m\}$  is a collection of  $m$  different items, given a transaction database  $D$ , where each transaction  $T$  is a collection of  $I$  projects, that  $T \subset I$ , and  $T$  has a unique identifier TID. If the item set  $X \subset I$  and  $X \subset T$ , then the transaction set  $T$  contains item set  $X$ . An association rule is an implication shaped like  $X \Rightarrow Y$ , in which  $X \subset I, Y \subset I, X \cap Y = \emptyset$ . The conditions for the establishment of association rules  $X \Rightarrow Y$ : a. It has the support of  $s$ , and there is at least  $s\%$  of the transaction contains  $X \cup Y$  in the database  $D$ . b. It has a confidence level  $c$ , that is,

in the transaction database  $D$ , there is at least  $c\%$  of the transaction that contains  $X$  also contains  $Y$  at the same time. The problem of association rule mining is to find the association rules between the minsup and the minconf which is given by user in the transaction database  $D$ .

Mining association rules can be roughly divided into two steps:

1) Find frequent item set from a collection of transactions;

2) Generating association rules from the frequent item set to satisfy the minimum confidence.

Here we use this kind of mining association rules based on data analysis methods for the entire network transformer defect data mining analysis.

## III. ASSOCIATION ANALYSIS OF TRANSFORMER DEFECT AND ANDEQUIPMENT PROPERTY

Defect rate is an important indicator of the potential risk assessment of equipment, and it is statistically obtained after the classification of the existing defects. But before proceeding with statistical analysis, it needs to sort out defect information in accordance with the relevant format for analysis.

Transformers defect data is collated well in accordance with certain format, then it can carry out the relevant mining analysis. Here we mainly analyze the association between device properties and defect rate of the grid transformer which put into operation in 2013 year, in which device attributes including manufacturers, voltage levels and put into operation in years, and the defect rate is calculated as follow:

$$P_{2013} = \frac{A}{B} \times 100\%$$

A = Corresponding transformer station defects times

B = Corresponding number of units put into operation transformer

Corresponding : manufacturers / voltagelevel / commissioning age

According to the manufacturer's defect statistics analysis, find out that the transformer defect rate in domestic manufacturers is generally higher than the foreign manufacturers, but the gap is not obvious, what's more the defect rate of part domestic manufacturers is significantly lower than foreign manufacturers. According to the statistical analysis of voltage grade defect rate, find out that the defect rate of 110kV and 220kV transformers is higher and similar, and 500kV transformers have lowest defect rate. According to statistical analysis of running years defect rate, find out that when equipment put into operation in early days (within 3

years) ,defect rate continued to rise; after the middle of the operation (4-13 years), defect rate is relatively stable; in equipment aging period (after 14) ,defect rate is rapid increasing, and these are in line with the basic defects tub curve. During the period of 11 years before operation, the possibility of the transformer leakage is in the 8% to 10% between, but in operation after 11 years, the possibility increased significantly more than 10%.

According to the order of device properties and defect rate, environmental factors and defect rate,property and equipment failure rate, environmental factors and the failure rate, each tap the potential risk factors of the equipment ,and then we can obtain a method for evaluating the potential risk state of a transformer. The analysis results are shown in the following table.

TABLE I. DEVICE PROPERTIES, ENVIRONMENTAL FACTORS INFLUENCE ON TRANSFORMER DEFECTS AND FAILURES.

Device properties	Influence of transformer defects	Influence of transformer failures
<b>Voltage level</b>	110kV and 220kV transformer defect rate is higher and similar, 500kV transformer defect rate is the lowest.	500kV transformer failure rate is the highest, the probability of potential defects develop to failure is maximum, therefore, recommended to focus on the operation and maintenance of 500kV transformer.
<b>Manufacturer</b>	The transformer defect rate in the domestic manufacturers is generally higher than the foreign manufacturers, but the gap is not obvious, and the defect rate of part domestic manufacturers is significantly lower than foreign manufacturers.	Family is obvious, the same manufacturer in the same batch of equipment containing the same design or workmanship defect.
<b>Running years</b>	When equipment put into operation in early days (within 3 years) ,defect rate continued to rise; after the middle of the operation (4-13 years) ,defect rate is relatively stable;in equipment aging period (14), defect rate is rapid increasing, and these are in line with the basic defects tub curve.	Transformer put into operation six years ago and put into operation 10 years ,its failure rate is high, and transformer run for more than 10 years,its failure rate is significantly increased.

**A. Relationship between Voltage Level and the Defect Rate**

According to the statistical analysis of voltage grade defect rate, finds out that the defect rate of 110kV and 220kV transformer is higher and similar, and the defect rate of 500kV transformer is the lowest. Statistics and calculation results of the corresponding transformer defect rate are shown in Table 2.

As can be seen, 220kV transformer has the highest defect rate, the defect rates of 110kV and 500kV transformers is lower

and similar; and conducted a statistical calculation for significant deficiencies, as shown in Table 3, it can be seen in major (emergency) defect chart that 110kV and 220kV transformer have higher and similar defect rates, 500kV transformer defect rate is lower, indicating the 500kV transformer’s manufacturing process was better than 110kV, 220kV transformers.

TABLE II. VOLTAGE LEVEL AND THE DEFECT RATE TABLE.

Voltage level	110kV	220kV	500kV	Total
<b>Total (unit)</b>	2060	551	223	2834
<b>Defective times (Times)</b>	903	437	77	1417
<b>Defect rate (times / unit)</b>	43.8%	79.0%	35.0%	50.0%

TABLE III. EQUIPMENT VOLTAGE LEVEL AND DEFECT (MAJOR)RATE TABLE.

Voltage level	110kV	220kV	500kV
<b>Total (unit)</b>	2059	551	223
<b>Defective times (Times)</b>	48	14	2
<b>Defect rate (times / unit)</b>	2.3%	2.5%	0.9%

**B. Relationship of Running Years and the Defect Rate**

According to statistical analysis of running years defect rate , finds out that when equipment put into operation in early days (within 3 years) ,defect rate continued to rise; after the middle of the operation (4-13 years), defect rate is relatively stable; in equipment aging period (after 14) ,defect rate is rapid increasing, and these are in line with the basic defects tub curve. During the period of 11 years before operation, the possibility of the transformer leakage is in the 8% to 10% between, but in operation after 11 years ,the possibility increased significantly more than 10%.

The relationship between the defect rate and running years Calculation method is that the transformer put into operation in 2013,its running years is 0, by analogy, calculate the running years of the transformer hang network operation of the province in 2013 and according to the corresponding number of defects times and operation of units calculated the corresponding defect rate , diagrams is such as Figure1. By the diagram shows, find out that when equipment put into operation in early days (within 3 years) ,defect rate continued to rise; after the middle of the operation (4-13 years), defect rate is relatively stable; in equipment aging period (after 14) ,defect rate is rapid increasing, and these are in line with thebasic defects tubcurve.

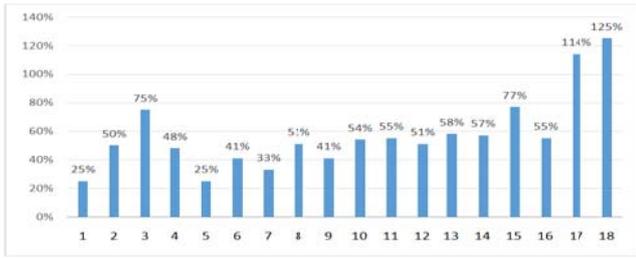


FIGURE I. THE RELATIONSHIP BETWEEN RUNNING YEARS AND THE DEFECT RATE.

The relationship between the leakage defect rate and running years

Transformer leakage defect and aging are closely related, have a statistical analysis of the relationship between leakage defect and the operating years, as shown in Table 4, figure 2. It can be seen that during the period of 11 years before operation, the possibility of transformer leakage is between 8% and 10%, but in operation after 11 years, the possibility is increased significantly more than 10% , so shall be given extra attention to the leakage problem on the transformer which put into operation over 11 years .

TABLE IV. LEAKAGE DEFECT RATE STATISTICS.

running years	Leakage defect units times	The number of units put into operation in that year	Defect rate
1	9	104	8.65%
2	10	152	6.58%
3	40	333	12.01%
4	31	377	8.22%
5	22	195	11.28%
6	11	162	6.79%
7	14	188	7.45%
8	23	202	11.39%
9	15	154	9.74%
10	24	164	14.63%
11	17	145	11.72%
12	18	97	18.56%
13	19	99	19.19%
14	9	60	15.00%
15	9	60	15.00%
16	9	42	21.43%
17	17	48	35.42%

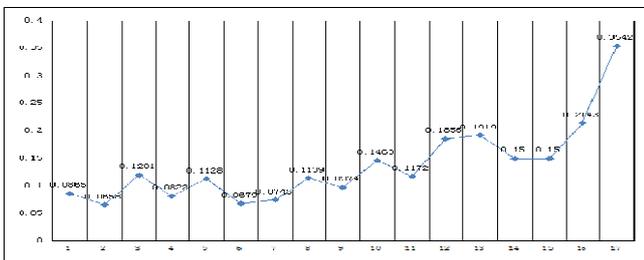


FIGURE II. THE RELATIONSHIPS BETWEEN LEAKAGE DEFECT RATE AND RUNNING YEARS.

The relationship between the defect rate of component failure and the running years.

The defect rate of the transformer component failure is

closely related to the running time. The statistical analysis of the relationship between the defect rate of component failure and the running years is carried out, as shown in Table 5 and figure 3. It can be seen that the first 10 years of operation, the defect rate of the component failure remained at a low level, after 10 years, the defect rate increased significantly.

TABLE V. COMPONENT FAILURE DEFECT STATISTICS.

operating years	Defect units times	Number of units put into operation	Defect rate (%)
1	0	104	0.00
2	1	152	0.66
3	1	333	0.30
4	4	377	1.06
5	2	195	1.03
6	0	162	0.00
7	0	188	0.00
8	2	202	0.99
9	3	154	1.95
10	1	164	0.61
11	1	145	0.69
12	1	97	1.03
13	2	99	2.02
14	6	60	10

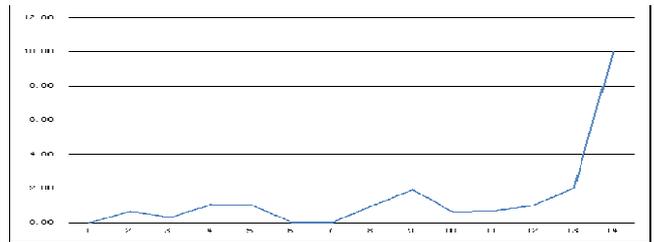


FIGURE III. THE CHANGE OF THE DEFECT RATE OF THE COMPONENT FAILURE AND THE RUNNING TIME.

#### IV. CORRELATION ANALYSIS BETWEEN TRANSFORMER DEFECTS AND ENVIRONMENTAL FACTORS

According to the order of device properties and defect rate, environmental factors and defect rate, property and equipment failure rate, environmental factors and the failure rate, each tap the potential risk factors of the equipment ,and then we can obtain a method for evaluating the potential risk state of a transformer. The analysis results are shown in the following table.

TABLE VI. DEVICE PROPERTIES, ENVIRONMENTAL FACTORS INFLUENCE ON TRANSFORMER DEFECTS AND FAILURES.

Environmental factors	Influence of transformer defects	Influence of transformer failures
Temperature	Negatively correlated with the leakage defect rate, defect rate is highest in winter and 110kV, 220kV transformer have higher leakage defect rate. Positively associated with the overheating defect rate, overheating defect rate is lower when the winter temperature is low, when the summer temperature is high, overheating defect rate is high .	In 2013,the entire network transformer failure concentrated happened in the higher temperature of summer, rarely occurs in the comfortable temperature of spring and autumn.
Precipitation	Positively correlated with the damp defect rate. With an increase in precipitation and increases, with a decrease in precipitation decreases.	Heavy rainfall can easily lead to poor sealing defect develop transformer internal damp and cause failures, the rain season has more failure-prone,
Lightning Level	With the heavy rainfall occurred simultaneously,threaten the communication equipment, control lines, etc..	Mainly result in a non-power transformer protection malfunction
Filthy	Because of the relatively good working environment in transformer substation, the influence is not obvious	Intensify local product pollution caused by uneven electric field distribution, have a greater threat in special equipment installation location .

A. Correlation Analysis of Temperature and Defect Rates

According to the statistical analysis of the environmental temperature and failure rate, finds out that in 2013,the entire network transformer failure concentrated happened in the higher temperature of summer. In 2013 the whole network failure of transformer and the temperature distribution is in Table 7, it can be seen in 2013,the entire network transformer failure concentrated happened in the higher temperature of summer, rarely occurs in the comfortable temperature of spring and autumn.

TABLE VII. TRANSFORMER FAILURE AND RUNNING YEARS DISTRIBUTION.

Month	Average temperature of 2013 year	Failure times	Failure rate
1	13.9	1	0.03%
2	15.2	0	0
3	18.1	0	0
4	22.4	0	0
5	25.8	0	0
6	27.8	0	0
7	28.9	1	0.03%
8	28.8	2	0.07%
9	27.5	0	0
10	24.7	0	0
11	20.1	0	0
12	15.5	0	0

B. Correlation Analysis of Precipitation and Defect Rates

According to the statistical analysis of precipitation and defect rate, finds out that in 2013, 75% of the whole network transformer failure occurs in the higher rainfall in July and August. Heavy rainfall can easily lead to poor sealing defect develop transformer internal damp and cause failures, the rain season has more failure-prone, it is proposed to strengthen the transformer maintenance checks in more rainfall season.

The precipitation and defect rate distribution in 2013 is shown in table 8. Can be seen that the high rainfall of July, August is the higher rate of transformer fault season. Therefore, it is suggested that the maintenance and inspection of the transformer is enhanced in the season of rain.

TABLE VIII. STATISTICAL OF PRECIPITATION AND FAILURE CONDITIONS.

Month	Precipitation (mm)	Failure times	Failure rate
1	8.8	1	0.03%
2	12.7	0	0
3	63.7	0	0
4	201.3	0	0
5	344.2	0	0
6	236	0	0
7	255.6	1	0.03%
8	390.5	2	0.07%
9	165.4	0	0
10	4.2	0	0
11	73.9	0	0
12	136.3	0	0

C. Correlation Analysis of Lightning and Defect Rate

According to the statistical analysis of lightning activity and defect rate, lightning and heavy precipitation events occur simultaneously, threaten communications equipment, control lines and other equipment, and will result in a non-power transformer protection malfunction.

In the 11 transformer failures occurred in 2011-2013, has 1 times that the 500kV main transformer, its vicinity of the line is struck by lightning, cause gas protection action. In 2013, the distribution of the lighting density and failure rate is shown in Table 9, and the transformer failure is more frequently happened in July and August. It is suggested that it needs to strengthen the maintenance and inspection of the transformer in the time and area of frequent lightning.

TABLE IX. LIGHTNING ACTIVITY AND FAILURE STATISTICS.

Month	Lightning density (time / square kilometer)	Transformer failure times	Failure rate
1	0.00591	1	0.03%
2	0.00997	0	0
3	0.87628	0	0
4	0.69936	0	0
5	2.76357	0	0
6	0.97344	0	0
7	0.94924	1	0.03%
8	4.07097	2	0.07%
9	0.98561	0	0
10	0.01011	0	0
11	0.0167	0	0
12	0.00995	0	0

#### D. Correlation Analysis Filthy and Defect Rate

According statistical analysis of filthy level and defect rate, finds our that the relationship between filthy index and power supply bureau failure rate is not obvious, but found that the power supply bureaus that failure rate is more than 1% are Q Power Supply Bureau, H Power Supply Bureau, O Power Supply Bureau and I Power Supply Bureau, its filthy index over 48.

In 2013 the pollution of the whole network transformer in the defect information is distinguished according to the pollution area level (A-E). In this case,  $a=100$ ,  $b=75$ ,  $c=50$ ,  $d=25$ ,  $e=0$  are calculated by using the numerical method to calculate the pollution area index. According to the data, the failure rate of the power supply bureau can be obtained, as shown in Table 10 and Figure 4. From the relationship diagram, the relationship between filthy index and power supply bureau failure rate is not obvious, but it is found that the power supply bureaus that the failure rate is over 1% are H power supply, O Power Supply Bureau and I Power Supply Bureau, the filthy index is more than 48.

TABLE X. FILTHY LEVEL AND THE FAILURE RATE STATISTICS.

Power Supply Bureau	Failure number	Runnin g units	Failure rate	Filthy Index
A Power Supply Bureau	0	54	0	58.3
B Power Supply Bureau	1	495	0.202%	44.7
C Power Supply Bureau	2	484	0.413%	62.3
D Power Supply Bureau	0	35	0	52.8
E Power Supply Bureau	0	216	0	34.4
F Power Supply Bureau	0	275	0	53.3
G Power Supply Bureau	0	88	0	44.2
H Power Supply Bureau	1	58	1.72%	54.8
I Power Supply Bureau	1	64	1.56%	48.3
J Power Supply Bureau	1	116	0.862%	55.2
K Power Supply Bureau	0	82	0	65.4
L Power Supply Bureau	0	46	0	58.5
M Power Supply Bureau	1	114	0.877%	26.4
N Power Supply Bureau	0	67	0	38.9
O Power Supply Bureau	1	61	1.64%	53.9
P Power Supply Bureau	0	119	0	17.7
Q Power Supply Bureau	2	106	1.86%	75.3
R Power Supply Bureau	0	218	0	54.4
S Power Supply Bureau	0	136	0	23.0

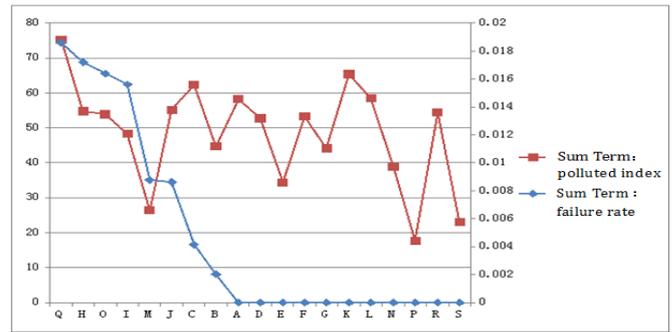


FIGURE IV. POWER SUPPLY BUREAU AND THE FAILURE RATE OF FILTHY INDEX A PLOT.

#### V. CONCLUSION

This paper presents an analysis method of transformer defect data based on association rules, and analyzes the association relationship among the defect data and equipment properties, equipment state and its own properties and the external environment.

By analyzing the relationship between the manufacturer, voltage levels and commissioning period and the failure rate obtain following conclusions:

1. There is a great relationship between the transformer failures and the manufacturing process of the manufacturer, which shows a strong familial characteristics, especially the transformer auxiliary equipment such as heavy gas relay, high pressure casing, etc. After the failure, it is necessary to check the equipment put into operation in same period.
2. Overall transformer failure was mainly due to manufacturing design flaws in the long run , and the higher the voltage level of equipment defects generate greater the impact.
3. Transformer put into operation six years ago and put into operation 10 years, its failure rate is higher, and transformer running more than 10 years, its failure rate was significantly increased.

Through the correlation analysis, let the original messy, huge data in tabulation, trend graphs way intuitively display to the user, the user can use these information after analyzing to have a reasonable and effective identification for transformer running conditions, and for future procurement of transformers and transformer repair and maintenance provided assistance, it has a strong guiding significance.

#### ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who helped me during the writing of this thesis. Without their consistent and illuminating instruction, this thesis could not have reached its present form.

#### REFERENCE

- [1] BRIDGMAN M S. Relating failure prognostics to system benefits [J].Aerospace Conference Proceedings, 2002 (7) : 3521-3526.
- [2] CHANG Shu-ping,WU Rui-tao. The Application of Failure Prognostic System in State Monitoring of Power Plant Generation Equipments [C]. GEESD, 2011 International Conference, Jilin : [s.n.],2011.

- [3] Zhengyuan Bing transformer fault characteristic confidence association rules [J]. High Voltage Engineering, 2012 (1).
- [4] Mo Juan, Wang Xue, Dong Ming, Zhang Yan. Power transformer fault diagnosis based on rough set theory, Chinese Society for Electrical Engineering, 2004, Vol. 24, No. 7, 162-167
- [5] Xiong Hao, Sun Caixin, Du Peng, comprehensive assessment of power transformer condition based on matter element theory [J] Chongqing University (Natural Science), 2006,29 (10): 24-28.
- [6] Zhang Wenjing, SONG Yu, Lu Haixia. Knowledge Association Rules Expert System [J]. Computer Technology and Development, 2006, 16 (2): 76-80.
- [7] Jia Guixia, Zhang Yong. Decision rule extraction method based on association patterns [J]. computer engineering and design, 2006, 16 (12):2175-2177.
- [8] Wu Qing, Cheng Haiming, Zhao Chunhua. Research on the association rules of fault diagnosis for oil monitoring [J]. Journal of Wuhan University of Technology: Transportation Science and engineering, 2004, 28 (5):103-105.
- [9] An Shuzhi. Warehouse and data mining [M]. Beijing: Tsinghua University press, 2005
- [10] Liu Tongming. Data mining technology and its application [M]. Beijing: National Defense Industry Press.2001.
- [11] Wang Huaqiu, Cao Chang Xiu, Wang Yue. A fast parallel algorithm and simulation research association rules hundred. Application of computer, 2004,21 (7): 223-225.
- [12] Li Jianxin, Zhang Qishan. Summary association rule mining algorithm [J]. China Engineering Science .2005.