Research on Intelligent Mining Algorithm for Distribution Network Transformer Fault Early Warning

J P Luo¹, L L Yuan², L Q Han² and W T Huang²,³

¹China Southern Power Grid, Guangzhou 510000, China
²Digital Grid Research Institute, China Southern Power Grid, Guangzhou 510000, China

E-mail: 280515123@qq.com

Abstract. The distribution network is the end link of the power grid. The level of management directly affects the power supply capacity and power quality. It is related to the operation level and social image of the power grid enterprise. The transformer is a very important component in the distribution network. The transformer is continuously safe and stable. Grid power supply reliability. In this context, this paper proposes an intelligent gateway intelligent fault detection algorithm for transformer operation failure, which collects data of transformers in multiple dimensions and multiple source channels, such as electrical variables and non-electrical variables. Correlation analysis and causality analysis of the data to achieve an all-round warning of transformer operation failure. The traditional method of transformer fault warning is solved by relying on the special gas concentration or transformer vibration signal generated by the fault of the transformer fault to solve the problem of fault identification and early warning singleness and low reliability, and fully exert the technical advantages of the Internet of Things and data mining. The evaluation value of the five "ancestor" elements is 0.2, 0.25, 0.15, 0.3, 0.1. Transformer operation fault trend early warning coefficient is set to 0.80. Once the early warning coefficient is found to exceed 0.80 during data collection and mining analysis, a transformer failure early warning will be issued, which can facilitate grid operators' intervention and control in advance. The actual calculation case also demonstrates the correctness and effectiveness of the intelligent mining algorithm. This paper provides an effective support for the development and application of big data technology in the distribution network.

1. Introduction
With the development and promotion of 5G technology, information technology and big data technology have gradually been applied and developed in the power grid. By applying big data technology to the distribution network, it is very important to improve the operation level and reliability of power supply networks. The use of big data technology to collect multi-source and multi-dimensional data in the power grid, collect a large amount of data and perform data identification and fault mining to achieve fault early warning is the focus of future power grid development.

Power transformers are the core components in the power grid. The operating status of the transformer is related to the safe and stable operation of the entire power grid. If the transformer operation process can be pre-warned and effective measures will be taken to troubleshoot the transformer before it fails, it will greatly reduce the transformer operation Failure rate [1]. There are many reasons for transformer failure, including electrical factors, including voltage level, current level,
load factor level, and power factor level; non-electrical factors include rainfall, humidity, main transformer temperature rise, ambient temperature, cooler failure, transformer sound, transformer life, main transformer insulation, transformer chromatogram, etc. Transformer fault early warning can be obtained by collecting the original data of the above factors of the transformer and identifying and digging it through algorithms.

In order to be able to effectively monitor and manage the operating status of the transformer, it is necessary to measure the dissolved gas content after the transformer heats up. Reference [2] uses the objective entropy weight method to identify and diagnose power transformer faults. The gas dissolution of 120 groups of power transformers is demonstrated in the power grid. According to the results of the demonstration, it can be known that the proposed method can overcome the problem of fuzzy coding in fault diagnosis.

Literature [3] is based on the transformer surface vibration signals to diagnose and identify faults. In the process of fault identification, the lamp vibration information of the winding is mainly related to the current, pretension, core vibration, and non-linear factors. The method of separating the winding vibration amplitude and the core vibration amplitude Differentiate and judge the state of the windings based on the vibration amplitude level of the load current windings, so as to identify and warn the transformer failure.

During the operation of the transformer, if a small amount of gas is generated due to excessive heating, the insulating material at the fault point will also accelerate to decompose, causing the gas content to exceed the limit. Literature [4] proposed an online early-warning method for the probability distribution of historical data of special gases, which was classified accurately. In the transformer state space, the transformer fault warning is issued through the identification and mining of historical data.

The above method only combines the heat generated by the fault point of the transformer to identify the special gas for fault early warning, or realizes the transformer fault early warning through the transformer temperature field. The historical data is limited, and it is difficult to collect a large number of multi-dimensional transformer data for comprehensive detection. [5-7]. The above-mentioned detection data are mainly non-electrical variables. It is difficult to effectively identify the causal relationship between electrical factors such as voltage level, power level, load rate and faults. Therefore, the comprehensive coverage of the data source was not truly achieved in the fault early warning [8-10]. Against this background, this paper proposes an intelligent mining early warning algorithm for transformer operation faults for the ubiquitous electric power Internet of Things. This algorithm collects data from transformers in multiple dimensions such as electrical variables, non-electrical variables, and multiple source channels. Data mining, and correlation analysis and causality analysis of the data, so as to achieve a full range of early warning of transformer operation failures.

2. Big data-oriented algorithm framework

First, various detector devices installed in the transformer device are used. The detector device detects voltage signals, current signals, temperature signals, gas concentration signals, humidity signals, transformer sound signals, transformer insulation signals, transformer chromatographic signals, vibration signals, and service life signals. Wait. Then collect multi-source, multi-dimensional, heterogeneous data, and optimize the data by optimizing communication protocols; use intelligent mining algorithms to perform correlation analysis and causality analysis on the data to establish a failure probability map model; according to the prediction, the possible failures will be caused Factors for actual application failure early warning and troubleshooting, big data-oriented algorithm flow system shown in Figure 1.
Figure 1. Process diagram for ubiquitous power internet of things algorithm

T test is Student t test, t is mainly used for the normal distribution with a small sample content and unknown overall standard deviation. The t-test uses the t-distribution theory to infer the probability of a difference, thereby comparing whether the difference between the two means is significant.

F test is joint hypotheses test, it is a test of statistical values subject to F-distribution under the null hypothesis. It is usually used to analyze a statistical model that uses more than one parameter to determine whether all or part of the parameters in the model are suitable for estimating the population.

After the multi-dimensional data of the transformer is collected, the correlation test is performed using the Pearson product moment correlation coefficient method, and the threshold is set for the T test. If it is less than the threshold, the correlation is not high, and the combination of factors can be filtered. For those larger than the threshold, it indicates that the correlation is high, and then perform Gran Causality detection to set the threshold. For F test less than the threshold, it indicates that the causal correlation is not high and eliminate the directed edge of the causal relationship. High degree of data is retained and a probability map model is established. After the probability map model is obtained, the transformer early warning analysis can be performed.

2.1. Correlation analysis

Different signal detection devices are installed in different locations and collect various electrical and non-electrical variables. Electrical and non-electrical variables constitute data set $D$. Data set $D$ is a multi-dimensional and time series variable set, $D = \{D_1, D_2, \cdots, D_M\}$, where $M$ is different types of data. The number of variable data sources, $D_i$ is any time-series data set in the data set.

The collected multi-source, multi-dimensional, heterogeneous transformer operating data has data redundancy or invalid data. In order to extract valid data with high correlation for data mining, it is necessary to perform correlation analysis on the collected data sets. Correlation analysis mainly refers
to statistical analysis of the original data through correlation algorithms, excluding data with low correlation and retaining data with high correlation. This paper uses Pearson product moment correlation algorithm for statistical analysis. The algorithm is as follows [11,12]:

$$R = \frac{\sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N} (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^{N} (Y_i - \bar{Y})^2}}$$

(1)

$X$ and $Y$ represent two data samples taken during the operation of the transformer, $X_i$ and $Y_i$ represent the values of the two samples at time $i$; $ar{X}$ and $ar{Y}$ represent the average of the two samples; $N$ represents the time series of the samples.

The larger the absolute value of $R$ is, the stronger the correlation between the two samples $X_i$ and $Y_i$ is. By judging the absolute value of $R$, the data set with higher correlation is retained.

2.2. Causal analysis

After the correlation analysis of the transformer operation data is completed, the causality analysis needs to be further completed. The causality analysis uses the Granger method to perform causality analysis on the two sets of variables $X_i$ and $Y_i$, using equations (2) and (3) China Granger causality assumes that transformer failure strong causality factor detection is performed, a transformer failure strong causality threshold is set, and data sets with analysis results greater than a predetermined threshold are retained[13,14].

$$Y_t = \sum_{i=1}^{T} \alpha_i X_i + \sum_{j=1}^{T} \beta_j Y_j + \mu_1$$

(2)

$$X_t = \sum_{i=1}^{T} \lambda_i X_i + \sum_{j=1}^{T} \delta_j Y_j + \mu_2$$

(3)

In the formula: $X_i$ and $X_t$ represent the values of data set $X$ at time $T$ and $i$; $Y_t$ and $Y_i$ represent the values of data set $Y$ at time $T$ and $i$; $\mu_1$ and $\mu_2$ are random white noises, and $\alpha_i, \beta_j, \lambda_i, \delta_j$ coefficient. If formula (2) is true, then it can be explained that $X$ is the cause of the change in $Y$, and if formula (3) is true, then $Y$ is the cause of the change in $X$.

Tested by Granger's method, a directed correlation matrix is established based on the test results to form a probabilistic model diagram.

In order to fully explore the "effect" and many "cause" factors, this paper uses a big data probability map model to analyze and analyze multiple factors affecting transformer failures, and specifically establishes a set of causal factors between faults and causes. Nodes represent random variables, and directed edges represent causal relationships between nodes. The causal relationship between nodes is demonstrated through the causal relationship between "parent nodes" and "child nodes". In the establishment of the probability graph model, the causal weight between node $X_i$ and parent node set $\pi(X_i)$ is recorded as $P(X_i \mid \pi(X_i))$, and formula (4) is causal Weight calculation formula.

$$P(X_i \mid \pi(X_i)) = P(X_i \mid X_1, X_2, \ldots, X_{i-1})$$

(4)
Correlation coefficient calculation method of fault or abnormal node \( X_i \) and multidimensional ancestor nodes in the causal structure of the probability graph in the transformer operation data. Specify that the faulty node \( X_i \) has \( m \) relatively independent ancestor nodes, and the ancestor node related to \( X_i \) is denoted by \( \pi_1(X_i), \pi_2(X_i), \ldots, \pi_m(X_i) \), \( \pi_j(X_i) \rightarrow X_i \) represents the causal relationship between \( X_i \) and the \( j \)th ancestor node, and the corresponding dependency coefficient \( P(X_i \mid \pi_j(X_i)) \) is denoted by \( \pi_j(X_i) \rightarrow X_i \). Weight \( \lambda^j_i \).

### 3. Intelligent mining algorithm and case analysis

The core of the intelligent mining algorithm defined in this paper is to collect a large amount of multi-source, multi-dimensional, and heterogeneous data. By integrating the Pearson product moment algorithm and the Gran Causality algorithm, a large amount of electrical variable data and non-electrical variable data are ensured. The relevance also ensures causality. The advantage of the intelligent mining algorithm proposed in this article is that it can process a large number of data from different sources and properties, including voltage data, current data, temperature data, gas concentration data, humidity data, sound data, insulation level data, chromatographic signal data, vibration data, Service life data, etc[15].

The intelligent mining algorithm system is set up as follows. The first step is to consider the collection of multi-source elements, and use the correlation algorithm to analyze the original data set \( D \), and retain the highly relevant data set \( E \). Then, perform a causal algorithm analysis on data set \( E \), Retain the data set \( F \) with strong causality, where \( F \subset E \subset D \). The comprehensive intelligent mining algorithm framework is shown in Figure 2.

![Figure 2. Comprehensive intelligent mining algorithm framework](image_url)

The intelligent mining algorithm includes three levels. The first level is data collection and classification. It collects multi-source, multi-dimensional, heterogeneous data of transformer operation, and collects voltage data, current data, temperature data, and gas concentration data. Humidity data, sound data, insulation level data, chromatographic signal data, vibration data, and service life data are sampled at time intervals to form data set \( X_1, X_2, X_3, \ldots, X_m \). The above data sets are corrected, filtered, and classified. According to different data categories \( A \ X_1, X_2, X_3, \ldots, X_m \)-time multi-dimensional sequence data set is formed, as shown in Part I of FIG. 2.

\( X_1, X_2, X_3, \ldots, X_m \) is used to perform correlation mining analysis according to formula (1), and the data larger than the threshold is retained to form a data set sequence \( \hat{E}_1, \hat{E}_2, \hat{E}_3, \ldots, \hat{E}_m \), as shown in part II in FIG. 3. In the data mining process, the correlation factor of Pearson product
moment is used, and the corresponding threshold is set to 0.7. The correlation detection result is shown in Figure 3.

![Correlation test result](image)

**Figure 3.** Transformer operation data for correlation analysis results

After data analysis, it is found that the factors such as "current", "temperature", "gas concentration", "insulation level", and "chromatographic signal" are highly correlated with each other.

Carry out causal relationship mining according to formula (2) or (3), and retain data larger than the threshold to form a data set sequence $F_1, F_2, F_3, \ldots, F_m$. Set the threshold of causality test to 0.05, and the data set to be analyzed for correlation is "high current", "current current", "current low", "temperature high", "temperature middle", "temperature low", "gas". The results of the causality test are shown in Fig. 4 as "high concentration", "medium gas concentration", "small gas concentration", "high insulation level", "medium insulation level", "low insulation level", "chromatographic signal".

![Causality test results](image)

**Figure 4.** Transformer operation data for correlation analysis results

The factors that can pass the Gran Causality test include "high current", "high temperature", "large gas concentration", "medium gas concentration", "low insulation level" and other factors that have a strong correlation with transformer failure.

According to the results obtained by the Granger causality test, a probability map model is constructed, as shown in Part III of Figure 2. The probability map model shows that there are five "ancestor" elements associated with the transformer failure ("fruit" factor), which are: Humidity, main transformer temperature rise, ambient temperature, transformer load factor, and cooler failure.

1) The probability of causing a transformer failure with a large current factor (1.1 times greater than the rated value) is 63.2%;
2) When the main transformer temperature rises (temperature is greater than 70 °C), the probability of causing transformer failure is 71.4%;
3) When the gas concentration is large (1.3 times greater than the limit value), the probability of causing transformer failure is 71.3%;
4) The probability of causing transformer failure when the gas concentration is greater than 1.0 times the limit value is 66.5%;
5) The probability of transformer failure at low insulation levels is 67.2%.

Based on the combination of causality analysis results and artificial experience judgments to form the final transformer failure decision model, intervention and elimination of the potential failure risk of the transformer are shown in formula (5).

\[ H(X_j, \pi(X_j)) = \sum_{j=1}^{n} a_j \pi_j P(\pi_j(X_j)) \]  (5)

In the formula: the actual operation expert evaluates the causal dependence of each group, combines the evaluation results and considers artificial experience to evaluate the coefficient \( a_j \) of the parameter structure rationality of \( \pi_j(X_j) \rightarrow X_j \), and \( \sum_{j=1}^{n} a_j = 1 \); \( P(\pi_j(X_j)) \) is the influencing factors during the operation of the transformer The set of real-time prediction probabilities.

Based on the results of the above analysis, a sensitive obstacle trend and early warning analysis result set is established. Calculate the real-time probability \( P(\pi_j(X_j)) \) corresponding to the ancestor node \( \pi_j(X_j) \) of the "fruit" node \( X_j \) that has a fault or has a fault tendency during the operation of the transformer, and set the threshold of the early warning coefficient \( H \) in Equation 5, and determine whether \( H \) is greater than the threshold, and if it is greater than , Then node \( X_j \) alarms, indicating that there is a risk of failure in the operation of the transformer, otherwise no alarm is issued.

The decision judgment model is constructed by combining expert experience judgment and the constructed probability map model, and the establishment of the early warning model through the adjustment of parameters, so as to determine whether a fault early warning is performed during the operation of the transformer. The evaluation value of the five "ancestor" elements is 0.2, 0.25, 0.15, 0.3, 0.1. Transformer operation fault trend early warning coefficient \( H(X_j) \) is set to 0.80. Once the early warning coefficient is found to exceed 0.80 during data collection and mining analysis, a transformer failure early warning will be issued, which can facilitate grid operators' intervention and control in advance.

4. Conclusions
Through actual data verification, the intelligent mining algorithm can simultaneously collect and identify a large number of electrical and non-electrical variables, and successfully warn the operation failure of the transformer. This fully shows that big data technology can make up for the defects and deficiencies of traditional data collection and identification methods that rely on limited variables such as gas concentration and oil temperature, which greatly improves the effectiveness of fault early warning. Can be known by actual putting into operation, this strategy can effectively improve the prediction accuracy rate by 25%, can reduce the occurrence of transformer failure by 45%. This solution has been gradually applied in the actual power grid and has achieved good results.

References
[1] Li G, Liu Y P and Fan H 2017 Challenges and prospects of fault prognostic and health management for power transformer Automation of Electric Power System 41(23):156-63
[2] Huang D R, Chen C S and Sun G X 2017 Recognition and diagnosis method of objective
entropy weight for power transformer Fault Automation of Electric Power System 41(12): 206-9
[3] Yan J K, Ma H Z and Li K 2017 Vibration signal based diagnosis method for looseness fault of transformer winding Automation of Electric Power System 41(3):122-6
[4] Zang L C and Li B 2015 Research on online warning based on probability distribution of historical data Yun Nan Electric Power 43(6): 48-50
[5] Li H, Wang F Z and Wang R, 2018 Identification algorithm for transformer insulation fault types based on improved RBF neural network Journal of Power Supply 16(5): 167-71
[6] Zhou K, Tu C M and Gao L K 2014 Faults and operation strategy of power electronic transformer Journal of Power Supply 6(11): 108-14
[7] Yan Y J, Sheng G H and Wang H 2016 The key state assessment method of power transmission equipment using big data analyzing model based on large dimensional random matrix Proceedings of the CSEE 36(2): 435-40
[8] Jiang Y W and Li Z W 2019 An information mining method of power transformer operation and maintenance texts based on deep semantic learning Proceedings of the CSEE 39(14): 4162-70
[9] Liao W H, Guo C X and Jin Y 2019 Oil-immersed Transformer fault diagnosis method based on four stage preprocessing and GBDT Power System Technology 43(6): 2195-202
[10] Wang Y Y, Zhou L W and Liang X H 2018 Markov forecasting model of power transformer fault based on association rules analysis Power System Technology 44(4): 1051-8
[11] Peng Z, Bo Qi and Zhi H R 2015 Differentiated warning of transformer based on data mining techniques Power System Technology 42(2): 951-3
[12] Yang Z and Shintemirov 2009 Association rule mining-based dissolved gas analysis for fault diagnosis of power transformers IEEE Transactions on Systems 36(6): 597-9
[13] Jin F, Feng W and Qiu S 2017 Hybrid RVM-ANFIS algorithm for transformer fault diagnosis Transmission and Distribution 11(14): 3637-43
[14] Yue W and Lin F 2019 Health status diagnosis of distribution transformers based on big data mining IEEE Trans Emerging Tel Tech. 12(11): 955-8
[15] La C, and Li Y 2018 Dissolved gas analysis principle based intelligent approaches to fault diagnosis and decision making for large Oil immersed power transformers IEEE Energies 8(8): 122-4