State Evaluation for Intelligent Distribution Terminal Units Based on Mining Association Rules

L Zhu¹, Y Ge¹, X D Song²*, W B Zhao³ and J C Guo⁴

¹Guangdong Power Grid Co. Ltd., Guangzhou, Guangdong 510000, China
²Guangdong Diankeyuan Energy Technology Co. Ltd., Guangzhou, Guangdong 510080, China
³Zhongshan Power Supply Bureau of Guangdong Power Grid Co. Ltd., Zhongshan, Guangdong 528400, China
⁴School of Electrical Engineering, Southeast University, Nanjing, Jiangsu 210096, China

*694011613@qq.com

Abstract. The construction of Ubiquitous Power Internet of Things expands the content and scale of data in distribution network. In order to meet the demand of intelligent operation and maintenance of distribution terminals, this paper proposes a method of mining association rules in data to realize evaluation of distribution terminal units. In this method, firstly, the Apriori algorithm with directional constraint of terminals’ functional module is used to mine the correlation indicators system under each divided functional module of terminals. Secondly, the comprehensive measurement of confidence and lift degree is used to increase the objectivity of evaluation’s weights of indicators which found by mining associations rules. Then, combined with the fuzzy comprehensive evaluation method, according to the correlation indicators system, the status evaluation of intelligent distribution terminals is realized. Finally, the historical status data of the actual disabled terminals are selected to verify the good effectiveness of the method that described in this paper.

1. Introduction
Distribution terminal units [1], as the important monitoring and control equipment in the process of intelligent power grid, have a direct impact on the reliability and security of the distribution network. There are many types of distribution terminal units, which are widely distributed in complex operating environment and dropped into abnormal state without perception in time [2]. The traditional planned maintenance methods, which are involved in study [3], are inefficient and difficult to meet the operation and maintenance requirements of distribution terminal units. Therefore, it is necessary to study effective methods of online status assessment and diagnosis of terminals, and to find out the abnormal running state timely and accurately, so as to avoid the deterioration of the state of terminals.

At the present stage, benefit by the development of the Ubiquitous Power Internet of Things [4], data collection, data mining and other technologies [3], in the distribution information system, the scale of data related to the operation of distribution equipment in grid continues to expand. In the context of big data, how to effectively collect, fuse, mine and utilize multi-source and heterogeneous data to realize the evaluation of power equipment and system state is a hot research direction at present.
Literature [5] evaluated the state of distribution terminals based on D-S evidence theory, but only selected simple statistics such as the off-line’s frequency of terminal equipment as the state characteristic. The selection of indicators was too conservative, which may affect the credibility of results of state evaluation. Literature [6] divides the multi-dimensional factors that affect the operation state of distribution network into main factors and secondary factors, and constructs a recursive model to evaluate the operation state of distribution network combined with analytic hierarchy process. However, the determination of indicators’ weights is solely dependent on a small number of experts’ judgement and mark, so how to determine the objective weights of data analysis is its improvement direction. Literature [5-6], all based on the comprehensive fuzzy evaluation method, respectively carried out improvement research on the state evaluation methods of mine hoist and relay protection. They all have good results in a certain test range, but the establishment process of state indicators lacks the process of data mining and analysis, which means the indicators is not established objective enough. Literature [7-8] quantifies the correlation between transmission lines’ characteristic factors and lines’ lightning strike faults with association rules, and establishes a lightning damage risk assessment model based on evidence theory, which is of reference significance for mining state data of distribution terminals' association rules.

This paper proposes a terminal state evaluation method based on data mining, aiming at the problems of complicated distribution terminal modes, hardware standardization objects that are difficult to establish state evaluation systems, difficult to determine state indicators under complex data, and the shortcomings of traditional evaluation methods, such as fixed mode and insufficient objectivity. First, the terminal equipment is divided into modules according to standardized functions, so as to establish the evaluation object of the device layer and the functional layer. Then, based on the Apriori algorithm with the constraints adding a single front and back item, the state indicators under each functional module of the terminal are mined respectively and objectively. Finally, by combining the comprehensive weight of each indicator with the objective item and the fuzzy state evaluation method, the terminals’ state evaluation completely based on the data is realized, and then the method will is verified with the reality example.

2. Partition and association rule mining of distribution terminal functional module

2.1. Functional module division of Distribution terminals

As complicated electronic equipment without standardized structure, distribution terminals’ equipment health and functional integrity are coupled with the equipment, internal and external factors. In order to improve the standardization and generalization level of terminals’ state evaluation process, the standardized functional structure and module division process of distribution terminals are firstly analyzed.

As the hardware structure and component composition of the terminals may vary depending on the model of the equipment and the manufacturer's process level, the diversity of the hardware structure will affect the standardized performance of the state evaluation method. Although there are hardware differences such as component structure and model of terminal equipment, the basic functions of its configuration have the requirements of standardization. The functional structure of the distribution terminal is divided into five modules, which are regarded as the second-order objects of terminals' state evaluation: CPU module, Power module, Communication module, Measure module and Control module, which are denoted as the set of terminals’ functional modules: \[ M = \{M_1, M_2, M_3, M_4, M_5\} \]

The running state of distribution terminal is divided into four evaluation elements, which constitute an evaluation set: \[ E_i = \{e_1, e_2, e_3, e_4\} \], representing the five functional modules of the distribution terminals in turn. \(e_1, e_2, e_3, e_4 \) respectively representing the normal, attention, abnormal and serious state of each functional structure.
2.2. Preprocessing of state parameters

The state parameters related to the operation of the distribution terminals are extracted from the internal operating parameters, external communication interaction and environmental state. The parameter names and labels determined in this paper are shown in Table 1.

The calculation formula of SOH (State of Health) of battery is shown in formula (1):

\[
SOH = \frac{C}{C_n}
\]

(1)

Where \( C_n \) is the rated capacity of the battery in the terminal; \( C \) is the current full charge of the battery.

| Parameter                        | Symbol | Parameter                        | Symbol | Parameter                        | Symbol |
|----------------------------------|--------|----------------------------------|--------|----------------------------------|--------|
| Time lag of GPS                  | \( z_1 \) | Secondary load of PT/CT          | \( z_2 \) | Insulation resistance            | \( z_3 \) |
| Secondary load of PT/CT          | \( z_4 \) | Signal strength                  | \( z_5 \) | The off-line rate                | \( z_6 \) |
| The number of maloperations      | \( z_7 \) | Battery’s SOH                    | \( z_8 \) | Period of message missing        | \( z_9 \) |
| Telemetry deviation              | \( z_{10} \) | Harmonic content                 | \( z_{11} \) | Number of message missing        | \( z_{12} \) |
| Time delay of communication      | \( z_{13} \) | Relative humidity                | \( z_{14} \) | Performance period               | \( z_{15} \) |
| Temperature                      | \( z_{16} \) | Humidity                         | \( z_{17} \) | A type terminal’s fault rate     | \( z_{18} \) |
| B type terminal’s fault rate     | \( z_{19} \) | C type terminal’s fault rate     | \( z_{20} \) | D type terminal’s fault rate     | \( z_{21} \) |
| The state of CPU module          | \( z_{22} \) | The state of measure module      | \( z_{23} \) | The state of control module      | \( z_{24} \) |
| The state of power module        | \( z_{25} \) | The state of communication module| \( z_{26} \) |

The statistical calculation process of terminals’ off-line rate is shown in formula (2). Its trigger is started by the total call message of the central station, which recorded every \( \Delta t \) minutes, and the statistic times of the normal response within the observation period (unit: day) are calculated as \( N_{\text{off}} \).

\[
Rate_{\text{off}} = 1 - \frac{N_{\text{off}} \times \Delta t}{24 \times 60 \times T}
\]

(2)

In addition, in order to explore the influence of common family defects in distribution terminal units on state evaluation, the fault rate of four types of distribution terminal in the observation range is taken as the parameter to be analyzed, and the fault frequency of the same type of terminal is calculated from the historical data of terminal operation and maintenance under the same statistical conditions.

When the status of each terminal module is abnormal or fault, the specific function missing module confirmed by the test will be marked as 1 in case of abnormal or fault, and 0 in case of normal. In addition, other state data without special description are all default to be general signal.

According to the above selected state data, it can be divided into numerical type and Boolean type, in which the numerical data can be divided into larger and better type, smaller and better type, and interval optimal type according to the response results. They all have their own dimensions. In order to facilitate calculation and analysis, they are firstly standardized according to the data characteristics. Then, in order to facilitate subsequent mining and processing of association rules between state
parameters, according to whether the difference between state parameters and standardized operating state is too large, the numerical parameter is changed to Boolean type, that is, 1 means significant deviation from standardized operating state, and 0 means within the tolerable variation range of parameter standards.

2.3. Association rules mining based on Apriori algorithm

With the improvement of smart grid information data collection and storage capacity, power distribution terminals have accumulated a large number of highly redundant parameter data in the process of operation and maintenance. The correlation between data parameters and terminals’ functional state can be analyzed according to the data mining method.

Association Rules are used to formalize the causal dependency relationship between the front items $X$ and back items $Y$, namely, it is expressed as $X \Rightarrow Y$, and then quantify the strength of the association relationship. Therefore the criterion of state evaluation of distribution terminal function module will be explored.

$Y_i$ is the back items of the association rules, which is used to indicate whether the functional modules of the distribution terminals are abnormal, with an outlier value of 1 and a normal value of 0. Each data parameter recorded and collected in the running process of the terminal is marked as $x_j$ ($j = 1, 2, \ldots, k$), then the possible combination of $k$ data parameters constitutes items $X_n$ ($n = 1, 2, \ldots, 2^k - 1$) as the front ones of association rules.

Set $D$ be the sample data set of parameters and functional module states, which accumulated in the historical operation and maintenance process of the distribution terminal. For mining association rules $X_n \Rightarrow Y_i$, generally, support and confidence degree are used to quantify the strength of association rules, and lift degree is used to judge the validity of association rules (The higher lift degree is, the higher correlation between the front items and back items is). In this paper, it is considered that the association rules with the degree of lift greater than 2.5 are valid, and their calculations are shown in equations (3), (4) and (5) respectively. In order to extract the association rules with high credibility as far as possible from the factors related to the state of terminals, the threshold of support degree was set as 0.01 in this paper, and the threshold of confidence degree was set as 0.4.

$$S(X_n \Rightarrow Y_i) = \frac{P(X_n \cup Y_i | D)}{num(D)}$$

$$C(X_n \Rightarrow Y_i) = \frac{P(Y_i | X_n)}{num(X_n)}$$

$$L(X_n \Rightarrow Y_i) = \frac{P(Y_i | X_n)}{P(Y_i)} = \frac{num(X_n \cup Y_i) \times num(D)}{num(X_n) \times num(Y_i)}$$

where $num(\cdot)$ represents the counting operator that relies on the sample data of terminals under the corresponding set of items.

In order to realize the state evaluation of distribution terminals’ modules, the Apriori algorithm adopted in this paper is constrained: the back item is limited to the state of a certain functional module of the distribution terminals, and the front item only considers the influence of single factor and does not explore the potential rules under their mutual combination, so as to reduce the search scope and time cost.

3. Fuzzy comprehensive evaluation of terminals’ functional module

The result of association rule analysis is used as the indicators system of state evaluation of terminal module $R_i = \{r_{ij} | j \in N\}$, where $r_{ij}$ represents the validity evaluation indicators mined by association
rules. Then, according to the fuzzy comprehensive evaluation methods, the status of the functional layer and the equipment layer of the terminal is evaluated.

3.1. The subjective and objective comprehensive weights of indicator sets

For the established indicator $r_{ij}$, in order to quantify the influence degree and level of the difference on the overall state of the terminal, it is often necessary to assign weight according to its importance. In the traditional weighting method, equipment experts generally judge the importance of each indicator based on experience, that is, the analytic hierarchy process quantifies the weight. The process of determining subjective weights $\omega^s_{ij}$ originates from the equipment operation and maintenance experience of the judgment. To some extent, it lacks objectivity and is difficult to implement.

The objective weight derived from the analysis of data samples is determined by using the performance indicator in the mining process of association rules, and its calculation formula is as follows:

$$\omega^o_{ij} = \frac{C_{ij} + L_{ij}}{\sum_{j=1}^{N}(C_{ij} + L_{ij})}$$  \hspace{1cm} (6)

Where $C_{ij}$ and $L_{ij}$ respectively represents the confidence and promotion degree corresponding to the indicators in the mining process of association rules.

Then, the calculation formula of subjective and objective comprehensive weight of indicator $r_{ij}$ is shown in equation (7). For the $M$ functional module state of the distribution terminal, the weight matrix of each indicator is $W = [\omega_1, \omega_2, ..., \omega_M]$.

$$\omega_{ij} = \frac{\omega^s_{ij} + \omega^o_{ij}}{2}$$  \hspace{1cm} (7)

3.2. Comprehensive fuzzy evaluation method

At present, the commonly used models of membership functions obey triangle distribution, half trapezoid distribution, S-type distribution, Gaussian distribution and etc. Considering that in the multi-indicators system, the construction and calculation process of triangle combined with half trapezoid membership function is relatively simple and accurate compared with the complex membership function, this paper chooses this method to calculate the membership degree of each indicator.

$$a_1(x) = \begin{cases} 1 & x \leq X_1 \\ \frac{X_2 - x}{X_2 - X_1} & X_1 < x \leq X_2 \\ 0 & x > X_2 \end{cases}$$  \hspace{1cm} (8)

$$a_2(x) = \begin{cases} \frac{x - X_1}{X_2 - X_1} & X_1 < x \leq X_2 \\ \frac{X_3 - x}{X_3 - X_2} & X_2 < x \leq X_3 \\ 0 & \text{else} \end{cases}$$  \hspace{1cm} (9)
Here, \( x \) represents the evaluation value of terminal deterioration according to the evaluation indicator, \( a_i(x) \) is the membership function under each state evaluation.

According to the above mentioned membership function, the fuzzy relation matrix of the corresponding terminals corresponding to each functional module can be written as \( (F_{M_i})_{4 \times 4} \), where the subscript \( M_i \) represents the functional module label of the intelligent terminal.

According to equation (12), the fuzzy comprehensive evaluation of distribution terminal functional layer is calculated.

\[
J_{M_i} = W_i F_{M_i}
\]  

(12)

Finally, the fuzzy comprehensive evaluation of the equipment layer, that is, the overall state of the terminal, is carried out by formula (13).

\[
J = \begin{bmatrix} J_{M_1} \\ J_{M_2} \\ \vdots \\ J_{M_5} \end{bmatrix} = \begin{bmatrix} \omega_{M_1} & \omega_{M_2} & \cdots & \omega_{M_5} \end{bmatrix} \begin{bmatrix} J_{M_1} \\ J_{M_2} \\ \vdots \\ J_{M_5} \end{bmatrix}
\]  

(13)

Where \( \omega_{M_i} \) is the weight of terminal functional module; Results according to the principle of maximum membership, then the status evaluation results of distribution terminal are finally determined.

4. Example
In a power distribution terminal operation tag samples collected by the data sets (a total of 600 data, covering 26 state of relative terminal), for mining association rules of terminal state evaluation, and analysis of one distribution terminal with a fault has occurred, the state data before the fault is to be obtained, on the state evaluation, to verify that the method described in this paper.

4.1. Comprehensive fuzzy evaluation method
According to the Apriori algorithm mentioned above, the scatter diagram of 18 association rules mined from the data set is shown in figure 1. The mined rules have high confidence and lift degree, which proves their validity. The pattern diagram is shown in figure 2, in which the size of the circle represents confidence degree and the color depth represents lift degree. It can be seen that according to the constraints in the paper, the mined rules refer to the status of terminal’s functional modules by the transaction front item, that is, the status evaluation factors of terminal’s functional modules are determined based on the data.
The mined results of association rules which shown in table 2, found that through the Apriori algorithm could not only mine general rules, such as $z_{12} \Rightarrow z_{26}$, which means high frequent of unresponsive message is related to the abnormal state of communication module, but also could mine special rules which are hard to find in usual ways, such as $z_{19} \Rightarrow z_{25}$, $z_{14} \Rightarrow z_{22}$ and $z_{14} \Rightarrow z_{24}$. The rule $z_{19} \Rightarrow z_{25}$ finds that when the fault rate of B type terminal units’ fault mostly happened in power module. The rules $z_{14} \Rightarrow z_{22}$ and $z_{14} \Rightarrow z_{24}$ find that the CPU module and control module are much more easy to drop into abnormal state when the relative humidity is rising high, which is well known that the high humidity environment might cause condensation at the electronic circuit part inside the terminal units. Then the subjective and objective comprehensive weight of each indicator is
determined by the method described in this paper. The evaluation’s indicator and weight of each functional layer of distribution terminal are shown in table 3.

Table 3. Terminal status indicators and weights table

| Equipment layer | Functional layer | Indicators | $\omega^1$ | $\omega^2$ | $\omega^3$ |
|-----------------|-----------------|------------|----------|----------|----------|
| Terminal units  | CPU module      | $z_1$      | 0.323    | 0.279    | 0.301    |
|                 |                 | $z_6$      | 0.441    | 0.454    | 0.448    |
|                 |                 | $z_{14}$   | 0.236    | 0.267    | 0.251    |
|                 | Measure module  | $z_2$      | 0.573    | 0.345    | 0.459    |
|                 |                 | $z_{10}$   | 0.326    | 0.377    | 0.352    |
|                 |                 | $z_{11}$   | 0.101    | 0.278    | 0.189    |
|                 | Control module  | $z_3$      | 0.327    | 0.352    | 0.340    |
|                 |                 | $z_7$      | 0.421    | 0.363    | 0.392    |
|                 |                 | $z_{14}$   | 0.252    | 0.285    | 0.268    |
|                 | Power module    | $z_4$      | 0.367    | 0.235    | 0.301    |
|                 |                 | $z_8$      | 0.301    | 0.241    | 0.271    |
|                 |                 | $z_{15}$   | 0.237    | 0.264    | 0.250    |
|                 |                 | $z_{19}$   | 0.095    | 0.260    | 0.178    |
|                 | Communication   | $z_1$      | 0.131    | 0.136    | 0.134    |
|                 | module          | $z_5$      | 0.167    | 0.221    | 0.194    |
|                 |                 | $z_9$      | 0.257    | 0.216    | 0.236    |
|                 |                 | $z_{12}$   | 0.241    | 0.200    | 0.220    |
|                 |                 | $z_{13}$   | 0.204    | 0.227    | 0.216    |

4.2. Stratified state evaluation of terminal units

According to the method described in this paper, the state evaluation of a distribution terminal is carried out, and the degradation degree of the terminal corresponding to each indicator is substituted into equation (8)-(11) to obtain the fuzzy relation matrix of each functional module, and then the fuzzy comprehensive evaluation result on the functional layer is obtained from equation (12) as follows:

Fuzzy relation matrix of module CPU module:

$$J_{M_1} = [0.0, 0.3753, 0.3831, 0.2416]$$

Fuzzy relation matrix of module measure module:

$$J_{M_2} = [0.5616, 0.3313, 0.1071, 0.0]$$

Fuzzy relation matrix of control module:

$$J_{M_3} = [0.6423, 0.2317, 0.1260, 0.0]$$

Fuzzy relation matrix of power module:

$$J_{M_4} = [0.2371, 0.3682, 0.3947, 0.0]$$

Fuzzy relation matrix of communication module:

$$J_{M_5} = [0.0, 0.2812, 0.5643, 0.1545]$$

Finally, the overall state evaluation result of the terminal on the device layer is obtained:

$$J = [0.1253, 0.2217, 0.3489, 0.3041]$$

Finally, the overall status evaluation result of the terminal on the equipment layer is obtained: that is, the result of status evaluation is found to be abnormal from the data of the terminal one week before the fault, which means it is in the transition period before the fault. The state history of the terminal indicates that the communication link intermittently initialized before the terminal was completely out of service. It proved that the state evaluation method in a certain data could indicate terminal's actual
operation state, and according to the results of state evaluation, it should as soon as possible to arrange the relative maintenance and replacement of the abnormal terminal, in order to shorten the time period without monitoring and measurement for distribution network influenced by the fault terminal.

5. Conclusion
This paper presents a state evaluation method of intelligent distribution terminals based on association rules. Firstly, the Apriori algorithm is used to mine the association rules of each functional module from the multi-dimensional state parameters and then establish the corresponding state evaluation’s indicators, using confident degree and lift degree to determine the objective weight of each indicator. Finally, combined with the fuzzy comprehensive evaluation, the state evaluation on the functional layer and the equipment layer of the terminal was carried out successively, and the following conclusions were obtained:

1) Compared with the general fuzzy comprehensive evaluation method, whose indicators establishment is completely derived from summary experience, the method proposed in this paper establishes the indicators system based on data analysis, which greatly increases the objectivity and validity of state evaluation method for terminals.

2) In addition, the indicators system can change according to the update and expansion of the data set, making the method more suitable for the situation of state evaluation of terminal equipment with fuzzy environment or high data quality.

Acknowledgments
This work was supported by Science & Technology Project of Guangdong Power Grid Co. Ltd. (GDKJXM20172867)

References
[1] Cong W, Sheng Y R, and Qi G F 2018 J. Automation of Electric Power Systems. 42(15) 77-85
[2] Zhang Z H, Zhou J, and Cai Y M 2017 J. Automation of Electric Power Systems. 41(13) 106-110
[3] Zhang J C, Chen L, and Zhang M Y 2019 J. High Voltage Engineering. 45(6) 1729-1736
[4] Zhang Y Q, Kou L F, and Sheng W X 2016 J. Power System Technology. 40(3) 768-773
[5] Li G, Zhang B, and Zhao W Q 2016 J. Automation of Electric Power Systems. 44(6) 31-35
[6] Ma Z, Zhou L M, Yuan H W 2019 J. Proceedings of the CSEE. 39(1) 130-140
[7] Jiang X C, and Sheng G H 2018 J. High Voltage Engineering. 44(4) 1041-1050
[8] Xie C Z, Bai J F, and Wang H B 2018 J. Proceedings of the CSEE. 38(21) 6233-6244