Study of Fault Diagnosis Distribution Network Based on Rough Set and Artificial Intelligence

Hui Zhou¹, Zhong Wang², Chunqing Shi¹, Chaoying Liu¹, ShiQin Zhao¹ and Ninghuan Zhang³*

¹Duyun Power Supply Bureau of Guizhou Power Grid Co., Ltd, Qiannan, Guizhou 558000, China
²College of Electrical Engineering, Sichuan University, Chengdu, Sichuan 610065, China
³Hangzhou Harmony Technology Co., Ltd, Hangzhou, Zhejiang 311121, China
*Corresponding author’s e-mail: zhangninghuan@yghm3.com

Abstract. The current real-time data collected by the power grid includes remote measurement, remote signaling, and other fault factors. The research in this paper is based on the fault diagnosis technology of rough set theory and self-learning theory, taking fault influencing factors as conditional attributes, fault type as decision-making attribute, and generating rough set rule table through self-learning reduction through a large number of fault history records, the influence of meteorological factors, environmental factors, equipment factors and other factors on the probability of equipment failure, which can realize the route Risk level prediction and fault location. Case analysis shows that the distribution network diagnosis technology based on rough set plays an extremely important role in predicting the risk level of each section of the line and positioning after a fault occurs.

1.Introduction
Rough set theory is a theory that reduces many influencing factors and obtains key influencing factors. It can quantitatively analyze inaccurate, inconsistent, and incomplete mathematical knowledge[1]. Rough set theory has been widely used in the field of fault diagnosis because of its strong fault tolerance, such as inverter open circuit diagnosis[2], transformer fault diagnosis[3][4], control circuit fault diagnosis[5]. Generally, rough set theory is often combined with other data processing theories to greatly improve the accuracy of diagnosis. Common ones include Bayes-rough set theory diagnosis model[6], rough set-genetic algorithm theory[7][8], evidence theory and rough set theory. In the field of power grid fault diagnosis, the application of rough set theory has achieved fruitful results. Tian Hailin optimized the rough method based on quantum genetic algorithm to greatly improve the speed and accuracy of power grid diagnosis[9], Zhang Chunge based on the consequences of power grid relay protection, based on Bayesian-rough set theory studies the great role of rough set in power grid diagnosis.

2.The background of Fault location

2.1. The data requirements of fault location
Based on the current situation of automatic operation and monitoring of distribution network lines, the
steady-state power big data analysis method can be used to identify and predict faults to achieve low-current grounding fault line selection and location. The data requirements are:

- Real-time collection of TTU power data of all distribution transformers on the 10KV power supply line of the distribution network (or the timeliness is stronger, the shorter the collection interval, the better).
- Real-time dispatching data of substation power supply lines (mainly power data)
- If there is a line segment, then real-time data collection of the segment location.
- Topological routing and line parameters of each 10KV power supply line in the distribution network.
- Data access for other influencing parameters (for example: weather, thunder and lightning, environment, geography, etc.).
- Each busbar of the substation has an open triangular PT secondary side voltage value measurement to realize the identification of the faulty phase line.

2.2. The judgment of fault

The traditional fault judgment method mainly uses the open triangle voltage to judge the fault phase. The criterion is shown in Table 1.

| Single-phase ground fault type | Various characteristics of ground voltage | Fault phase discrimination | Open triangle voltage value (V) and phenomenon |
|-------------------------------|----------------------------------------|-----------------------------|-----------------------------------------------|
| Single-phase fully grounded   | The voltage of 1 phase is decreased but not zero, the voltage of two phases is increased but not equal, and the voltage of 1 phase may be slightly higher than the line voltage | The phase with zero voltage is the ground phase | 100V Voltage stability |
| Single-phase incomplete grounding | The voltage of one phase increases and does not exceed the line voltage, and the voltages of the other two phases decrease, but are not equal. | The next phase of the voltage increase phase is the ground phase (according to the positive phase sequence: A-B-C-A shall prevail) | <100V Voltage indication is unstable |

The traditional fault identification method has played an important role in the identification of the fault phase and the fault line. However, due to the fault judgment based on the outlet PT voltage of the circuit breaker, the specific location of the fault and the cause of the fault cannot be determined.

2.3. Failure factors

Refer to the ground fault statistics table provided by Duyun Bureau in the past two years (more than 90 grounding and more than 300 trips), try to extract effective information and calculate the fault risk based on human experience, roughly dividing the fault factors that may cause the fault into Meteorological conditions, environmental conditions of the equipment, factors of the equipment itself, and additional factors at the time of failure.

- Meteorological conditions refer to the specific weather and meteorological records at the time of the failure, including: weather (clear, cloudy, rain, snow, etc.), wind level, rain level, snow level, lightning level, icing level, fog level, and temperature.
- Environmental conditions of the equipment, including factors such as the equipment is located in a commercial area and far away from towns
- The equipment's own factors, including: whether the base station is dedicated line,
equipment aging degree, equipment quality, etc.

Additional factors when the fault occurs, including: month of occurrence, time of occurrence, type of bus grounding (or some other electrical characteristics)

According to the statistics of ground faults, the types of faults are classified and counted, which can be roughly divided into the following types of faults: damage to the arrester, broken wire and rod, tree pressing, damaged porcelain vase, man-made activities, loose and falling wires, equipment failure, and drain wire burnt Breaking, contact with foreign objects, causing damage, PT burst, switch damage, and cable burnout.

3. Aided analysis of rough set theory

3.1. Knowledge expression system

A knowledge expression system S can be represented by a quadruple S=(U,A,V,f), where U is the universe (a set of objects U={x1,x2,x3,...,xn}); A is the set of attributes (features, variables), V is the set of attribute values, Va is the value set of attribute a, which also becomes the value set of a; f: UxA information functions, defined for each a and A An information function f(x,a), that is, the information function f specifies the attribute value of each object x. The attribute set includes condition attribute value and decision attribute set.

Take the statistically-obtained risk factors as the condition attributes of the rough set and the fault type as the decision attributes of the rough set. From this, the fault history records are extracted to generate a decision table, and the attribute reduction of the condition attributes is performed, and finally the distribution network grounding is obtained Failure risk assessment model for failure. When future failures occur, input the conditions at the time of the failure into the condition attributes of the risk assessment model, so as to obtain the possibility of what type of failure occurs under the current situation.

3.2. Self-learning reduction of decision table

Through the decision table reduction method, the fault characteristics of the distribution network are refined, so that various additional factors are used as fault classification conditions and related attributes are established. The original data information is simplified to obtain the minimum reduction set. The self-learning reduction process of the reduction model is as follows:

Construct the initial decision table based on historical fault samples. According to the topological structure of the distribution network, the existing fault information is screened, a training sample set is established, and the decision-making information is determined as the fault location and fault type, and the condition attribute is the risk factor obtained by the above statistics.

1) Create the initial attribute matrix AD=(Aij)mXn of the decision table;
2) If the single attribute element is contained in matrix A, then put it into set K, if set K is a core attribute set, skip to step (4), otherwise, the disjunction logic of each non-zero element in the initial attribute matrix The expression is Lij=ixai, where ai is the attribute item in the non-zero element Aij, go to step (5);
3) For any Ki∈K (i=1,2,...,n), if there is any Ki∈AD, let the element Aij in AD corresponding to Ki=0 to obtain a new matrix A′D For all the non-zero elements in the matrix A′D, calculate the corresponding disjunctive logical expression Lij=ixai, where a, is the attribute item in the non-zero element Aij;
4) Take all the disjunctive expressions Lij conjunctive operation to obtain the conjunctive normal form L=ꓥLij
5) Convert the conjunctive normal form L into the form of the disjunctive normal form to obtainL’=IpL
6) Incorporate the elements in the core attribute set K into each conjunct item in L’, then each conjunct item corresponds to the result of an attribute reduction, and obtains the reduction set R, which is the experience value set of the failure cause. Defined as a static coefficient of failure probability.
7) When a distribution line fault occurs, collect and identify the known potential factors and summarize them to form a potential value set of the current fault cause, and calculate the proportion of such fault factors based on their static coefficients, which is defined as the dynamic coefficient \( R \) of the failure probability.

8) After weighting the dynamic coefficient \( R' \) and static coefficient \( R \) of the failure probability, the proportions of various failure causes are re-calculated, defined as the derivation coefficient of the failure probability, and the values are sorted from large to small, which is important for troubleshooting. According to the basis, guide the operation and maintenance personnel to find the line fault.

9) When the distribution line fault is checked and confirmed, the fault classification summary table of the area is updated, and the proportion of various fault causes is calculated, and the static coefficient \( R \) of the failure probability is updated.

Through the prior failure probability reduction and the self-correction of the fault information system, the accuracy of the fault location system is greatly improved.

4. Analysis of typical cases
There are seven measuring points A, B, C, D, E, F, and G on a certain 10kV power supply line. The failure probability set is shown in Table 2.

| Line section | Condition attributes | Decision attributes (P) |
|--------------|----------------------|-------------------------|
|              | Weather condition \((P_m)\) | Equipment environment \((P_k)\) | Equipment factor \((P_e)\) | Other factors \((P_o)\) |
| [A,B]        | 0.4556               | 0.3221                  | 0.2041                  | 0.0182                  | Broken wire |
|              | 0.0001               | 0.5214                  | 0.2145                  | 0.2604                  | Artificial activity |
|              | 0.3214               | 0.0001                  | 0.6211                  | 0.0574                  | Lightning arrester failure |
| [C,D]        | 0.2368               | 0.62                    | 0.0523                  | 0.0959                  | Broken wire |
| [F,G]        | 0.003                | 0.002                   | 0.8520                  | 0.143                   | Transformer damage |

This set is an initial decision table calculated based on historical records, where section AB passes through residential areas, and there is a tall building close to the line, and the lightning arrester at point A has been used for a long time. The CD section passes through the forest area under the mountain, and the transformer at the end of the FG section is old. The system assigns functions and colors according to different environments and different parts of the line. The highest levels of various risks are marked as different colors. Above 0.8 is red, 0.5-0.8 is orange, 0.3-0.5 is yellow, 0.1-0.3 is light green, and 0-0.1 is green.

In the decision table calculated based on historical records, the [C, D] section is only affected by other factors, and the highest probability of failure is

\[
P = P_m Z_m + P_k Z_k + P_e Z_e + P_o Z_o (Z_m = Z_k = Z_e = 0, Z_o = 1) = 0.0959
\]

The segment is shown in green on the topology map. After line inspection, it is found that there is a tree at point H 50 meters away from point C in the CD section and the distance from the line is less than 2 meters, then the probability of wire breakage at point H is

\[
P = P_m Z_m + P_k Z_k + P_e Z_e + P_o Z_o (Z_m = Z_k = Z_e = 0, Z_o = 1) = 0.7159
\]

Point H is marked in orange. According to the local weather forecast, if a thunderstorm occurs on the line from 15 to 18 o’clock on a certain day, the probability of disconnection at point H after the weather forecast is issued is

\[
P = P_m Z_m + P_k Z_k + P_e Z_e + P_o Z_o (Z_e = 0, Z_m = Z_k = Z_o = 1) = 0.9527
\]

The H point is marked in red, and a warning is issued that there is a risk of disconnection of the CD section.

The transformer at point G is old but cannot be replaced in time due to various reasons, and its risk level is always marked in red. And it is known that the transformer is prone to failure when the electric
load changes drastically. On the day of the failure, the remote signal showed that the circuit breakers on the low-voltage side of the transformer at point G were all disconnected; remote measurement showed that the load at the transformer outlet at point G had a sudden change of 30%, and the system judged that the transformer fault caused a power outage. After multi-party coordination, the transformer at point G was replaced, and the problem of transformer outage due to the state of health was solved. The probability set of transformer damage at point FG became \{0.43, 0.29, 0.25, 0.03\}, and the probability of occurrence was \(P=P_m*Z_m+P_k*Z_k+P_e*Z_e+P_o*Z_o(Z_0=1, Z_m=Z_k=Z_e=0)=0.03\). However, due to construction in the FG section, the risk probability set caused by human activities is \{0.234, 0.23, 0.01, 0.526\}, and the probability of failure is \(P=P_m*Z_m+P_k*Z_k+P_e*Z_e+P_o*Z_o(Z_m=1, Z_k=Z_e=0) = 0.23\), is the highest risk level of the segment, and the topology diagram is displayed in light green.

5. Summary

This article explains the background of fault location, including the current data of the distribution network, traditional fault judgment methods, and the factors that may lead to the distribution network and its data access. Based on this, a fault risk and fault location algorithm based on rough set theory is proposed. The algorithm combines rough set and self-learning to organize the original fault records into rough set inference rules through rough set theory, and enter the fault information after the fault occurs. Afterwards, the learning is carried out to improve the accuracy of the distribution network fault risk prediction and fault identification.

Reference

[1] Zhang K, Yang J. (2017) Overview of rough set theory and its applications. J. Internet of Things Technology, 7(06): 93-94+98.
[2] Han S.M, Zheng S.Q, He Y.S. (2020) Open-circuit fault diagnosis of inverter based on rough set greedy algorithm. J. Power System Protection and Control, 48(17): 122-130.
[3] Bian L, Liu C.Y, He H, Sun H.N, Dai J.G. (2020) The application of longhorn whisker algorithm in fault diagnosis of oil-immersed transformer. J. Application Research of Computers, 37(S1): 155-156.
[4] Zhou G.Y, Ma S.L. (2020) Research on Transformer Fault Diagnosis and Location Based on Machine Learning and DGA. J. High Voltage Electrical Apparatus, 56(06): 262-268.
[5] Liu J, Jiang Q.H, Liu L, Wang D, Huang D, Deng L.J, Zhou Q.C. Research on the resistance change type fault diagnosis and wind speed sensor location optimization of mine ventilation system. J/OL. Journal of China Coal Society: 1-8
[6] Xia C.H, Hu S, Li L.L, Liu Y.F, Cheng S. (2020) Application of diagnosis model and method based on Bayes-rough set theory. J. Journal of China Three Gorges University (Natural Science Edition), 42(05): 82-87
[7] Guo W. (2018) Application of Rough Set-Genetic Algorithm in Fault Diagnosis of Hydroelectric Generating Unit. J. Science and Technology Vision, 35: 144-146+154.
[8] Jin B. (2019) Compensation capacitor fault analysis method based on rough set and genetic algorithm. J. Shandong Industrial Technology, 04:167.
[9] Tian H.L, Hong L, Wang Y.X, Wang X.H. (2018) Power grid fault diagnosis based on quantum genetic algorithm optimization of rough-Petri net. J. Journal of Xi'an Institute of Technology, 32(06): 678-684.