ATLANTIS
PRESS

Prediction of Line Fault Based on Optimized Decision Tree

Jianping Yang[2], Jingxian Qi[1,*], Kang Ye[2] and Youlin Hu[2]
[1]Nari Group Corporation State Grid Electric Power Research Institute, Nanjing, China
[2]State Grid Shanghai Electric Power Company, Shanghai, China
[*]Corresponding author

Abstract—**Power line is the main equipment of transmission link, and its failure prediction will help to improve the safe operation level of the power grid and ensure the reliable electricity consumption of users. In this paper, the optimized decision tree algorithm is used to forecast fault line tripping. Firstly, correlation analysis algorithm of line run data attribute is being used to mining associate rules to build the analysis object. Secondly, mutual information is used to confirm the influence degree of various factors on the line tripping. Finally, based on the optimized decision tree algorithm to build the line fault prediction model, realize the early warning analysis of the line tripping. Through the implementation of the above methods, the active perception ability of the power grid is further enhanced to ensure the reliable operation of the power grid.**

Keywords—mutual information; decision tree; line fault; correlation analysis

I. INTRODUCTION

With the expansion of the power grid, the quantity of power equipment is increasing. And lines as the main transmission link for equipment, its failure will directly influence users' electricity consumption, so the decision tree algorithm is adopted to build the line fault prediction model, and it provides the line fault accurately forecast theory and method of support.

At present, there are genetic algorithms, integrated detection method, signal source method and bayesian network algorithm. The genetic algorithm [1,2,3,4] is mainly realized by fault location in the fault prediction and diagnosis of transmission lines, but this algorithm is less adaptable to the network operating mode changing at any time. The comprehensive detection method [7] makes fault prediction based on current and voltage changes, and it is necessary to consider various possible situations to avoid the invalidation of the judgment. The signal source method [7] needs to install the signal source cabinet in the substation, which requires coordination between multiple departments, and it is not easy to operate and maintain. Bayesian network algorithm [1] requires parameter setting based on domain knowledge in the construction of line fault prediction model, which is difficult.

Aiming at the above problems, this paper puts forward two suggestions for improvement of line fault prediction model: 1) improve the operability of model practice; 2) introduce more information to model construction. First in this paper, an improved decision tree algorithm to the fault line and the non-fault line reference attribute correlation analysis, avoid introducing the repeatability of information, and information theory of mutual information index is used to confirm the influence degree of all kinds of information of fault happened, finally construct the decision tree line fault prediction, through the fault early warning and enhance the level of safe operation of power grid.

II. MODELING

Transmission lines are the basic components of the power grid. The distribution range is wide, the operating environment is complex, the fault types of the lines and the causes of line faults are varied. The fault duration can be divided into transient fault and permanent fault. According to the cause of the problem, can be divided into circuit grounding, short circuit, circuit overload, because with the device itself (such as wire broken stocks, injury and flashover, etc.), external force damage fault and natural causes. For fault prediction of transmission lines, build the object model in this paper for the running line, which includes the name of the line, line running defects, line maintenance record, line fault tripping, deformation situation of switch, protection action information, remote sensing, weather information, thunder and lightning monitoring information, the changes of power transformer on-line monitoring information, the model built is as follows:

(sb_id,is_gz,is_jx,alarm_yx_kg,alaram_yx_bh,alarm_yc_dl, sbd_dxfb,sbd_dxwd,sbd_wf,sbd_wh,sbd_dxhc,sbd_gtqx,sbd_dldl,qx_fs,qx_wd,qx_sd,qx_qy,qx_jy,is_ld), where sb_id is the line ID and is the unique identifier of the line; is_gz represents the line is fault or not, is_jx represents the line is in maintenance state or not, alarm_yx_kg represents having switch displacement or not, alaram_yx_bh represents there is any protection action or not, alarm_yc_dl represents there is a huge change on behalf of the line current or not, sbd_dxfb represents line conductor ice coating, sbd_dxwd represents line conductor temperature, sbd_wf represents vibration line whether there is a breeze, sbd_wh filth on behalf of the lines, sbd_dxhc represent wire sag, sbd_gtqx represents line tower inclination , sbd_dldl represents the current of the line cable sheath, qx_fs represents meteorological information of wind speed, qx_wd represents the temperature of the meteorological information, humidity qx_sd represents meteorological information, qx_qy represents meteorological information of air pressure, qx_jy represents meteorological information of rainfall, is_ld represents the presence of lightning. Then, the data of line fault operation data and non-fault operation data are

collected, and each type of information is divided hierarchically, and then the sample data is analyzed based on the above lines.

III. DECISION TREE ALGORITHM

Decision tree based on line fault and the fault history operation data for calculation of fault classification and prediction algorithm, is a representative of object properties and a mapping relationship between object values, in which each node represents an object, each branch path in the tree represents a possible attribute values, and each leaf node is experienced from the root node to the leaf node path represents the value of the object. This algorithm analyzes the historical data with definite results, and finds the characteristics in the data, and then predicts the new data results. This article constructed their theoretical basis for using ID3 decision tree algorithm, the algorithm based on information theory, the mutual information of the highest value is chosen as the decision point, step by step to confirm may result in line fault factors.

A. Information Theory

Information is a universal form of the state of objective things and the characteristics of motion, which exist, produce and transmit various kinds of information in these forms in the objective world. Information theory is a goalkeeper information as the research object, on the basis of the characteristics and law reveals the essence of information, the application of mathematical method to study the information storage, transmission, processing, control and use of general rules such as science [8]. The basic concepts include information entropy, conditional entropy and mutual information.

1) Information entropy:

For a symbol P in a source, when the probability of the symbol is p, its uncertainty is defined as:

$$\mathrm{f}(P) = \log \frac{1}{p} = -\log p \qquad (1)$$

The information entropy describes the average uncertainty of the source, and its calculation formula is as follows:

$$E(T) = \sum_{i=1}^{c} -p_i \log p_i \qquad (2)$$

In which C represents the possible category of the source of the source, and p_i represents the probability of the occurrence of class.

2) Conditional entropy:

Set X, T as two discrete random variables, the variable X is given, then conditional entropy of the variable T is $E(T \mid X)$,which represents the uncertainty of variable T under the condition of known variables X, the calculation formula is:

$$E(T \mid X) = \sum_{i,j} p(x_i, t_j) \log \frac{p(t_j)}{p(x_i, t_j)} \qquad (3)$$

Where i, j is the value range of variable X and T, and p(*t*) represents the probability of $T = t$, p(*x*,*t*) represents the probability of the simultaneous occurrence of variable X and variable T.

Conditional entropy can eliminate the uncertainty in the unitary model by obtaining more information, which is to reduce the entropy of the unitary model through binary or multivariate models.

3) Mutual information:

For two variables X and T, if the information entropy of a variable X for E (X), the variable T information entropy for E (T), the mutual information G (T, X) said variable X contains information about the variable T, its computation formula is:

$$G(T, X) = E(T) - E(T \mid X) = \sum_{i,j} p(x_i, t_j) \log \frac{p(x_i, t_j)}{p(x_i)p(t_j)} \qquad (4)$$

Where G(T,X) represents the correlation between the variable T and the variable X, i.e. the degree of uncertainty of the variable T in the given case of variable X.

Mutual information is used to measure the correlation between information. When two messages are completely relevant, the mutual information is 1 and 0 is irrelevant.

B. Build Decision Tree

In this paper, ID3 algorithm is used to construct the decision tree. The ID3 algorithm is a method of constructing decision tree based on Occam's razor principle, which measures the selection of attributes by mutual information. Mutual information was used to measure information correlation between indicators, ID3 algorithm through computing the historical data of each category or attribute information entropy to gain mutual information, and select the mutual information of the highest category or attributes as the decision node in decision tree, the category or the value of the attribute as a branch or continue to split. Repeat this process until a complete decision tree is generated.

IV. OPTIMIZED DECISION TREE

Aiming at the shortcoming of decision tree algorithm, which ignores the correlation between attributes in data set, we use Apriori algorithm to analyze the correlation of object attributes, and retain only one attribute with strong correlation, thus avoiding repeated reference of attributes. That is, the optimized decision tree algorithm first uses correlation analysis to confirm the attributes of the analysis objects, then uses mutual information index to confirm the influence degree of each attribute to the line failure. Finally, the ID3 algorithm is used to build the decision tree.

A. Correlation Analysis

In this paper, the Apriori algorithm is used to analyze the property of line objects. Apriori algorithm is a basic algorithm for mining frequent item sets needed to generate association rules. This algorithm uses the prior knowledge of frequent itemsets, uses iterated search iterations, and K itemsets are used to explore k+1 itemsets. It is defined as:

itemsets: a collection of data items;

k-itemset: An item set containing K data items;

Support degree:for a set of A, B, the support degree represents $A \bigcup B$ ratio containing itemsets in transaction D. The calculation formula is as follows:

$$support(A => B) = p(A \bigcup B) \qquad (5)$$

Confidence degree: the confidence of the association rules is that the transaction containing the item set A also contains the proportion of the item set B. The calculation formula is as follows:

$$confidence(A => B) = p(B \mid A) = \frac{support(A \bigcup B)}{\sup port(A)} \quad (6)$$

The process of mining strong association rules is to confirm the association rules that satisfy the minimum support threshold and minimum confidence threshold. The process of the algorithm is as follows:

First, find frequent 1- itemsets, and record them as L_1. Then we use L_1 to mine L_2, namely frequent 2- itemsets, and keep circulating until more frequent k- itemsets cannot be found.

B. Build Optimized Decision Tree

First, association analysis of modeling object attributes is done to confirm the correlation between attributes. Then, unrelated attributes are selected as object attribute set for building decision tree, and decision tree is built based on ID3 algorithm. The flow chart of the algorithm is as follows:
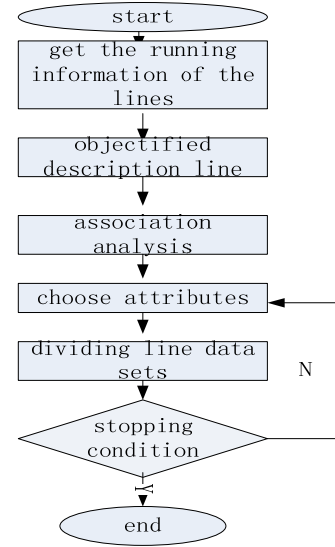

FIGURE I. DECISION TREE PROCEDURE

The termination conditions of the algorithm are divided into two cases: 1) all attributes belong to the same class; 2) no attributes can be used for decision making.

V. EXPERMENTAL RESULTS

Based on historical line fault records and running data of a certain area from 2011 to 2017, in order to adapt to the decision tree algorithm, the classification of each attribute is as follows:

TABLE I. CLASSIFICATION OF EACH ATTRIBUTE

Data Type	Classification	Unit
line defect	critical,serious,general,no	
line fault	tripping, single-phase grounding, overloading, and self reasons	
line maintenance	Y,N	
protection atcion	Y,N	
switch displacement	Y,N	
current change	Y,N	
icing thickness of wire	(0,10) ,[10,20) ,[20, ∞)	mm
wire temperature	(0,70],(70,90],(90,130),(130, ∞)	℃
breeze vibration	(0,0.5) ,[0.5,10) ,[10, ∞)	m/s
SPS	Y,N	
conductor sag	(- ∞ ,-2%),[-2%,4%) ,[4%, ∞)	
pole leaning	(0,0.5%) ,[0.5%,1.5%) ,[1.5%,5%)	
Power Cables current	(0,1) ,[1,100) ,[100, 200)	A
humidity	(0,60%) ,[60%,80%) , [80%,100%)	
temperature	(- ∞ ,-10],(-10,30],(30, ∞)	℃
pressure	(0,1000) ,[1000,1020) ,[1020, ∞)	Pa
wind speed	(0,3],(3,9),(9,12)	stage
thunder	(- ∞ ,-10) ,[-10,10) ,[10, ∞)	

First, based on the model construction, we collect the data, including line fault, line defect, line maintenance, remote alarm, telemetry alarm, breeze vibration, tower tilt, cable sheath

current, temperature and wind speed, and the specific models are as follows:

(sb_id,is_gz,is_jx,alarm_yx_kg,alaram_yx_bh,alarm_yc_dl, alarm_yc_yg,sbd_wf,sbd_gtqx,sbd_dldl,qx_fs,qx_wd) ;

Secondly, the Apriori algorithm is used to analyze the attributes of the reference. The support degree is 50%, the confidence level is 95%, and no association rules are generated. Therefore, all the attributes are selected to build the decision tree.

Finally, a decision tree is built based on the ID3 algorithm for the collected data. The results are as follows:
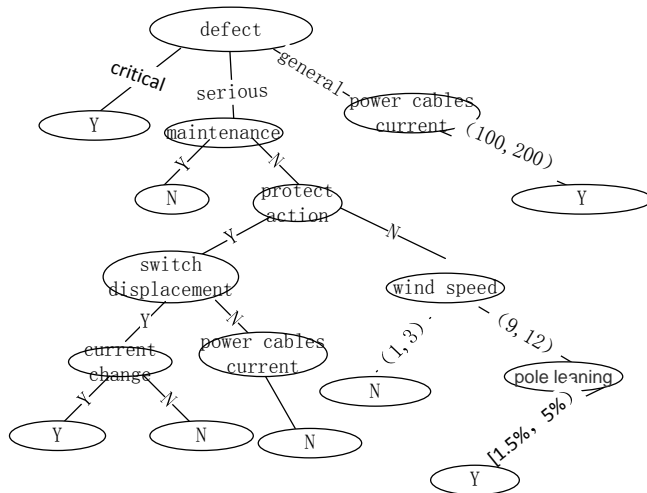

FIGURE II. EXPERIMENTAL RESULTS GRAPH

It is known from the decision tree that the line defect has an important influence on the failure. If there is a critical defect, then will the trip accident happened; if serious defects and defects, according to the existing switch shift, relay remote alarm, according to the change of the telemetry data can confirm the trip accident will happen; at the same time, if the weather is windy, tower inclination reaches 5% [1.5%.), will also trip accident.

The selection of threshold for association analysis is based on expert experience. At the same time, data set can be set up, and its rationality can be confirmed by analyzing association rules.

VI. CONCLUSION

In this paper, a line fault prediction method based on mutual information and improved decision tree is proposed, and the traditional decision tree algorithm is improved by association analysis algorithm, avoiding repeated reference of strong correlation attributes. At the same time, through the introduction of information entropy, conditional entropy and mutual information, construct a decision tree for the statistics provides theoretical support; in addition, this method has high flexibility for the construction of the object, can increase or decrease according to the actual object attribute data can be introduced. The example shows that the improved method

proposed in this paper can predict the fault of the line, and further ensure the safe and reliable operation of the power grid.

REFERENCES

[1] Liang C, Xiaoxuan W,Fault prediction and diagnosis of transmission line based on genetic algorithm,Digital Technology and Application, 2013 (1) :108-108
[2] Weijian R, fault diagnosis of power system based on genetic algorithm,2006
[3] Ping M, Optimization of neural network based on genetic algorithm and its application in fault diagnosis, chinese sciential and technical theses,2009
[4] Ruijin L, expert system of transformer insulation fault diagnosis and fault prediction based on genetic algorithm,2003(06)
[5] Zhengmo R,Cunkai Y, Yi z, Prevention and method of grounding fault of 10kV line, Major science and technology,2004(28)
[6] Haiqiang Z, failure prediction and analysis of 10kV distribution line are discussed, China High Technology Enterprises, 2013(5):127-128
[7] Agyemang M, Ezeife CI. Lsc-mine: Algorithm for mining local outliers. Proc. of the 15th Information Resource Management Association (IRMA) International Conference. New Orleans. 2004, 1: 5–8
[8] Gang L, Power transmission line tripping analysis of 35kV and above, China power education,2010
[9] Tao S, Wangling H, Baoquan W, influence of humidity on the measurement of frequency electric field of high voltage transmission line, High voltage technology, 2014,40(6)
[10] Hao Z,Yuhong Y, Discussion on some important issues in the development of uhv transmission in China. grid technology, 2005,29(12):1-9
[11] Liqiu Y,Haihua L,Yuzu G, hazard analysis of wind vibration of overhead transmission lines, Chinese science and technology information,2008(12)
[12] Xiaozhi W,Jian H,Guangxin Z,Meng R, A clustering decision tree framework algorithm in data flow classification. Computer security,2014
[13] Hongyu G,Qing C,Bingyin X, New algorithm for single - end wave fault location of transmission line. Power system automation. 2017,41(5):121-127
[14] Zhirui L, Shengsuo N,Nan J, Current situation and development trend of ac transmission line parameters measurement. 2017,41(11):181-191
[15] Jian L,Yichao C,Hong J, Rule-based general expert knowledge base fault diagnosis method. Computer and digital engineering. 2010,38(6)