XLPE cable health assessment based on Relief-F feature weighted FSVM

Xiaohuan Wu¹, Yang Liu², Liqun Wang¹,³, Xinzhuo Ren¹ and Xianjun Tan¹

¹ Hangzhou Power Equipment Manufacturing Co., Ltd., Hangzhou 310018, China
² Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education(Northeast Electric Power University), Jilin 132012, China
³ Corresponding author’s e-mail: 1593339758@qq.com

Abstract. For the cross-linked polyethylene (XLPE) cable health status assessment process, there are noise data for each feature parameter and the importance of each dimension feature is difficult to determine, which leads to deviations in the cable status assessment results. An XLPE cable health assessment algorithm based on Relief-F feature weighted fuzzy support vector machine is proposed. The Relief-F algorithm is used to calculate the importance of each dimension feature. The feature quantity that is weakly related to the evaluation of the health status of the XLPE cable is deleted. Furthermore, the fuzzy support vector machine is used to evaluate the health status of the XLPE cable, which gives the noise data a smaller degree of membership and reduces the influence of the noise data on the evaluation results. Experimental analysis shows that this method can accurately assess the health status of XLPE cables. It can provide a reference value for relevant departments to carry out cable maintenance and replacement work and save the human resources, and material resources required for maintenance and replacement of cables.

1. Introduction

XLPE Cable, as the critical power transmission equipment of the power system, plays an indispensable role in the safety of power load and reliable power transmission [1]. When the cable is in poor health, it is prone to fire or line faults caused by the fusing of the neutral wire, which seriously affects production safety quality. Therefore, timely and accurate understanding of the health level of XLPE cables is the technical support that managers must have for the operation and maintenance of the wires and for making replacement decisions.

For assessing the health status of XLPE cables, there have been many related studies in recent years. Some developed countries have conducted a lot of research in the field of XLPE cable insulation online monitoring and developed online monitoring methods such as the dielectric loss factor method and partial discharge method, and the developed online monitoring devices are widely used in cable insulation monitoring. A large amount of measured data has been accumulated, which provides a data basis for data-driven cable health assessment [2-4]. However, a cable health assessment pair with a single feature cannot fully reflect the health of the cable. Through the analytic hierarchy process, expert experience, and other methods, combined with the characteristics of each health status of the cable [5], the corresponding cable diagnosis or status evaluation model can be established to evaluate the health status of the cable [6]. The experimental results show that the
evaluation results of the cable modeling can be consistent with the actual operating state. Still, the impact of the noise data during the monitoring process on the evaluation of the cable health status is not considered.

In order to solve the above problems, this paper proposes to apply the classification algorithm of Relief-F feature weighted fuzzy support vector machine to the health assessment of cables. The Relief-F algorithm is used to calculate the weight of each feature part for evaluating the health status of the XLPE cable, and the feature quantities that are weakly related to the cable health evaluation are deleted. The weighted Euclidean distance from the sample to the center of the class is calculated, and then the membership function is constructed and classified based on the weighted Euclidean distance. On the one hand, the algorithm uses fuzzy support vector machines to reduce the impact of noise data on state evaluation. On the other hand, it considers the effects of feature importance on the classification effect and reasonably generate the membership degree between each dimension feature and state quantity, to accurately evaluate XLPE Cable health status.

2. XLPE cable health parameters

During the operation of the cable, as the health status of the cable changes, the status parameters also change. Choosing appropriate status parameters can improve the accuracy of cable status evaluation. XLPE power cable status parameters are divided into internal factors and external factors according to the factors that affect its operation. The specific cable health parameters are shown in Figure 1.

![Figure 1](image)

**Figure 1.** Status parameters that affect cable health.

Each state quantity can directly or indirectly reflect the health state of the XLPE cable [7]. To accurately evaluate and analyze the health of the cable, this paper selects the partial discharge reflecting the insulation performance of the cable, the cable corrosion reflecting the mechanical performance, the operating life reflecting the historical record and the annual load reflecting the influence of external factors as the characteristic parameters for the health of the cable. The status is evaluated.

2.1. The amount of partial discharge

Partial Discharge (PD) refers to a non-penetrating discharge phenomenon that occurs under the action of an electric field when an insulating material has an insulation defect. The size of the partial
discharge can be obtained by the ultra-high frequency detection method, the pulse current detection method and the acoustic wave detection method. At present, international authoritative organizations such as IEC, IEEE and CIGRE recommend partial discharge test as the best method for evaluating the insulation state of power cables, and it has been widely used in actual projects. The size of the partial discharge indicates the quality of the insulation performance of the power cable and whether there are safety defects in the running cable.

2.2. **Cable corrosion**
Because power cables are placed underground for a long time, they will inevitably be eroded by soil, moisture and other factors. After the cable has been in operation for a certain number of years, the power cable will show corrosion, especially the metal parts will show different degrees of corrosion. The degree of corrosion determines the health of the cable to a certain extent. The degree of corrosion can be qualitatively observed through visual inspection and microscopic morphology observation. It can also be quantitatively analyzed through elemental component analysis.

2.3. **Operating life and load conditions**
Under normal circumstances, after XLPE is put into operation, it will be subjected to various stresses such as electricity, heat, mechanics, and moisture, which will cause aging. The longer the operating life of the cable, the greater its health will be affected by uncontrollable factors. The service life can reflect the aging of the cable to a certain extent. Different load conditions also have a coupling phenomenon to the health of the cable. Under normal circumstances, the load of the cable can also be generally divided into three types: light load, heavy load, and overload. The overload circuit has large current carrying capacity, high cable operating temperature and fast aging speed. In order to describe the characteristic parameters of the load condition, the peak value of the current carrying capacity during the operation of the cable is used to describe the load condition of the cable.

3. **XLPE cable health status classification algorithm based on status parameters**
In order to reduce the influence of noise data and non-important features on the evaluation of the cable health status, this paper uses the Relief-F feature-weighted fuzzy support vector machine classification algorithm [8] to evaluate the cable health status, which can better reflect that the sample belongs to the correct category. The degree of membership calculation based on Relief-F feature weighting accurately gives the noise data a smaller membership degree, thereby reducing the impact of noise data on classification. The overall framework of the method flow is shown in Figure 2.

![Figure 2. Diagram of method flow.](image-url)
3.1. Feature selection based on Relief-F algorithm

The Relief-F algorithm used in this paper is a feature selection algorithm based on the Relief feature selection algorithm that supports multiple classification problems. This method calculates the weight of each feature, selects the elements strongly related to the health of the XLPE cable, and uses the importance of the part to establish the membership function of the fuzzy support vector machine. The influence of different characteristic values on the classification effect of the classifier under the running state of XLPE cable is fully considered.

There is a total data set $J$, the total number of features is $c$, and the total number of categories is $t$. Suppose the feature weight value $W$ of each dimension feature is initialized to 0, and the weight value will be updated every time the sample is sampled. Set the sampling times of the model as $m$ and the number of nearest neighbor samples as $r$. Let $F_{k,b}$ denote the $b$-th selection of the $k$-th non-same sample set. After sampling, the feature weight of the $j$-th feature is:

$$W_j = W_j - \sum_{a=1}^{r} \frac{\text{diff}(T_j, x, H_a)}{(mr)} + \sum_{k \neq y} \left[ \frac{P_k}{1 - P_y} \times \sum_{b=1}^{r} \frac{\text{diff}(T_j, x, F_{k,b})}{(mr)} \right]$$

Where $\text{diff}(T_j, x, H_a)$ is the distance between the sample $x$ and the $a$-th sample $H_a$ of the same sample set on the $f$th element; $h$ is the distance between the sample $c$ and $g$ on the $f$th feature; $P_k$ represents the $k$th sample in the training site $D$; $P_y$ represents the probability that the class of piece $x$ appears in the training set $D$. After calculating the weight of each feature, sort the importance of each segment from large to small, set a threshold $\Gamma$, and accumulate the feature weights, knowing that the accumulated value exceeds $\Gamma$ times the sum of the feature weights. Delete the features corresponding to the remaining feature weights to complete the feature selection.

3.2. Calculation of weighted euclidean distance

Suppose that after feature selection of the Relief-F algorithm, the feature weight set of the remaining $c'$ features is $W'$. The training set after feature selection is denoted as $D'$, and the $k$-th training set is represented as $D_k$ ($k = 1, 2, \cdots, t$). In order to find the weighted Euclidean distance from the sample to the class center, the mean point of all the pieces in $D_k$ on the $n$ features should be obtained and used as the class center $O_k$ of the $k$-th class. After receiving $O_k$, calculate the weighted Euclidean distance from a particular sample $x_i$ of $D_k$ to $O_k$ of the class center using the calculation formula:

$$d_{w,i,O_k} = \sqrt{\sum_{j=1}^{c'} W'_j (x_{ij} - O_{kj})^2}$$

Where $d_{w,i,O_k}$ is the weighted Euclidean distance from the sample $x_i$ to the class center $O_k$; $W'_j$ is the weight of the $j$($j=1,2,\cdots,c'$) feature; $x_{ij}$ is the component of the model $x_i$ on the $j$th feature; $O_{kj}$ is the component of the class center $O_k$ on the $j$th feature.

3.3. Design of membership function

Based on calculating the weighted Euclidean distance, a membership function based on the weighted Euclidean distance is designed. Let the radius of the hypersphere of the $k$-the training set be denoted as $d_{w,\text{max}}$. For any sample $x_i$ in the entire training set $D'$, where $y_i$ is the category to which $x_i$ belongs, its membership degree can be calculated by the following function:

$$\phi_i = 1 - \frac{d_{w,i,O_k}}{d_{w,\text{max}} + \delta}$$

(3)
Among them, \( d_{u, o}^{w} \) represents the weighted Euclidean distance from the sample \( x_i \) to the center of the \( y_i \) type; \( d_{y_{\text{max}}} \) represents the hypersphere radius of the \( y_i \) type training set; \( \delta \) is a small positive number set in advance to ensure \( 0 < \delta \leq 1 \). Through calculation, the corresponding membership degree of each sample in the entire training set can be obtained, thereby obtaining fuzzy training samples.

According to the above description, the total number of features in the complete data set of the original multi-classification problem can be reduced, the total number of categories remains unchanged, and the corresponding degree of membership is calculated for each sample because the fuzzy support vector machine is a two-class model. Therefore, for multi-classification problems, a one-to-one approach is used to train multiple binary classification models. Finally, the voting mechanism is used to output the category with the highest number of votes as a result.

4. Case analysis

4.1. Collection of experimental data and cable health status comments

The data set used in the experiment comes from the Kaggle website [9], which contains the health status assessment data of 2500 XLPE cable sections in a particular place in western Canada in 2003, 2008, 2013, and 2018. It includes partial discharge value, cable corrosion, and peak data of service life and current carrying capacity during operation. During the experiment, 8000 groups in the data set were randomly selected as the training set according to the number of health status data in Table 1, and the other 2000 groups were used as the test set. In the selected data set, 5000 noise data is randomly added to the training sample, and 1000 noise data is randomly added to the test sample. The remaining unchanged data is clean data. The data contained in the data set is shown in Table 1.

According to Q/GDW456-2010 "Guidelines for the Evaluation of Cable Line Status", four types of XLPE cable health status comment sets are defined, namely "normal status", "attention status", "abnormal status" and "serious status", namely \( V = \{\text{Normal, Attention, Abnormal, Serious}\} \). The state quantities of the normal state are at the warning value specified in the regulations; note that the state reliability decreases and the change trend of the state quantities is are approaching the standard limit, but does not exceed the standard limit; the state quantities of the abnormal state have changed greatly, close to or slightly Exceed the standard limit; Severe state: The quantity of each state seriously exceeds the standard limit.

| Health status classification | Training samples/piece | Test sample/piece | Status label |
|-----------------------------|------------------------|------------------|--------------|
| Normal status               | 2700                   | 600              | \( V_1 \)    |
| Attention status            | 1800                   | 400              | \( V_2 \)    |
| Abnormal state              | 1800                   | 600              | \( V_3 \)    |
| Severe state                | 1700                   | 400              | \( V_4 \)    |

4.2. Analysis of results

Since Relief-F is used to calculate the weight of each feature, it is necessary to set a threshold \( \Gamma \) to select the part in the process of feature selection. The selection of \( \Gamma \) value significantly affects the classification effect of the classifier. Therefore, this article explores the accuracy of the classifier under the threshold of 70%, 80%, 90%, and 100%. To avoid errors in the experimental process, the experiment was repeated ten times, and the average of the accuracy of the classifier under each threshold was taken to analyze the experimental results. The experimental results are shown in the figure below:
Figure 3. Experimental results under different thresholds.

As shown in Figure 3. According to the experimental results, it can be found that when the $\Gamma$ value is 90%, the accuracy of the classifier is the highest, reaching 98.54%. At this time, according to the calculation results, the remaining characteristics are the amount of partial discharge, the degree of cable corrosion, and the current carrying capacity. The useful life is deleted.

To verify that this method improves the accuracy of the cable health assessment by reducing the number of features and giving the noise data a smaller degree of membership. The fuzzy support vector machine method (RFSVM) based on Relief-F feature weighting, the standard SVM method and the standard fuzzy support vector machine (FSVM) method are compared for experiments. In order to reduce the influence of random disturbance on the experimental results, various classification method experiments were repeated 10 times. The three methods evaluate the cable health status as a set of tests, and the training set and test set data used in each set of tests are the same. Each method adopts the cross-validation method to select the kernel function and penalty factor with the best classification effect. The experimental comparison results are as follows:

Figure 4. Comparative experiment results.

As shown in Figure 4, the average accuracy of the fuzzy support vector machine method based on Relief-F feature weighting in XLPE cable health assessment reached 98.54%. It is higher than the standard Euclidean FSVM method, with an average classification accuracy of 95.9%. And the standard SVM method with an average classification accuracy of 93.98%. Experiments have proved that the RFSVM method, which assigns reasonable membership to noise data and non-important features, can obtain a high-accuracy XLPE cable health evaluation result. The validity of this method in the assessment of cable health status is verified.
5. Conclusions
This article focuses on the existence of noise data in the XLPE cable health status data set, and the difficulty of determining the importance of each dimension feature leads to errors in the health status assessment. Therefore, this paper uses Relief-F feature weighted fuzzy support vector machine classification algorithm to evaluate the health of XLPE cables. The experimental results show that this method can more accurately judge the health status of XLPE cables, which provides a reference for the maintenance and replacement of XLPE cables. In the follow-up work, with the large-scale promotion of cable online monitoring devices, more strongly correlated characteristic parameters are introduced. Improve the accuracy of the algorithm proposed in this paper to evaluate the cable health status.

Acknowledgment
This work was supported by Hangzhou Power Equipment Manufacturing Co., Ltd. The project number is YF200501.

References
[1] Yu J, Chen X, Meng F, et al. 2020 Numerical analysis of thermo-electric field for AC XLPE cables with different service times in DC operation based on conduction current measurement [J] IEEE Transactions on Dielectrics and Electrical Insulation 27(3) 900-908
[2] Zhang Z, Assala P D S, Wu L 2018 Residual life assessment of 110kV XLPE cable [J] Electric Power Systems Research
[3] Su Y, Liu Y, Zhong L 2019 Evaluation of voltage endurance characteristics for new and aged XLPE cable insulation by electrical treeing test [J] IEEE Transactions on Dielectrics and Electrical Insulation 26(1) 72-80
[4] Sayidul Morsalin, Animesh Sahoo, Bao Toan Phung 2019 Recovery voltage response of XLPE cables based on polariation and depolarisation current measurements [J] IET Generation Transmission & Distribution 13(24)
[5] Wang Yaqun, Yin Yi, Li Xuguang, Cai Chuan 2009 Isothermal relaxation current used in the life assessment method of 10kV XLPE cable[J] Transactions of the Chinese Society of Electrical Engineering 24(09) 33-37+52
[6] Li Dengshu, Wang Xin, Zhao Ming, Huang Xiaowei, Lou Yufeng 2020 XLPE insulation aging state assessment based on set pair analysis dynamic weighting method [J] Hydropower Energy Science 38(07) 194-197+172
[7] Su Y, Liu Y, Zhong L 2019 Evaluation of voltage endurance characteristics for new and aged XLPE cable insulation by electrical treeing test [J] IEEE Transactions on Dielectrics and Electrical Insulation 26(1) 72-80
[8] Ye G 2018 Simulation of Electrical Trees in XLPE Cable Insulation and Electric Field Analysis [J] Journal of Engineering
[9] https://www.kaggle.com/utilityanalytics/utility-underground-cable-dataset1