Electromagnetic Interference Diagnosis Based on HPLC Timing Sequence Topology and Machine Learning

Haojie Tian¹, Yudong Zhao¹, Liyan Kang², Ying Shang² and Xinran Liu²

¹ Liaoning Electric Power Co., Ltd., Shenyang, Liaoning, 110006, China
² Electric Power Research Institute of Liaoning Electric Power Co., Ltd., Shenyang, Liaoning, 110006, China

Abstract: High power line communication (HPLC) has become a basic communication method in the courts of smart grid areas. Some special electrical equipment in these power grid areas may cause electromagnetic interference, which will undermine the stability of HPLC, and lead the acquisition equipment in the smart grid to miss out or lose control. Consequently, the power supply quality and grid safety will be affected. This kind of influence is temporary, random, and self-recoverable, and it is difficult to be identified and eliminated through on-site inspection. In this paper, the HPLC network topology records of the courts are analyzed, the HPLC timing sequence topology is fitted, and user information and electricity consumption information are correlated. Machine learning and optimization are carried out on the data of the courts by using the association rule learning and decision tree algorithm. It is found that the machine learning method based on the HPLC timing sequence topology can effectively realize the electromagnetic interference positioning of the power grid in the courts, and help the on-site operation and maintenance personnel to conduct troubleshooting.

1. Introduction

As the HPLC technology becomes more and more mature, the anti-interference ability of HPLC is gradually strengthened, but there are still occasional electromagnetic interferences from the electrical equipment. Due to the openness of the low-voltage power grid, the access of some new equipment or counterfeit equipment will cause harmonics that intrude into the low-voltage power grid causes electromagnetic interference. Electromagnetic interference also has a serious impact on power line carrier communication, which can affect the collection and communication of nearby users, and even cause abnormal communication in all stations[1].

This paper mainly uses the machine learning method to study the network topology information of the smart grid based on the location diagnosis of electromagnetic interference[2], and then the correlation between electromagnetic interference and network topology of courts is analyzed. So as to realize the electromagnetic interference source location and then check the fault, and find the electromagnetic pollution, purify the grid environment, improve the quality of power supply, and ensure the safety of electricity supply.

2. Background

Currently, HPLC communication is becoming more and more popular in low-voltage power supply networks. The smart grid information system can receive and record the topology information of the power grid at a time. The smart grid information system can also initiate acquisition commands and collect network topology information of the courts according to requirements[3]. HPLC has the ability...
of adaptive networking and route optimization. After the network communication node is faulty or strongly interfered, the network route is automatically adjusted to form a new network topology. And when the load of courts communication network is different, the topological expressions will be different in different time periods.[4]. The ideal routing link is close to the physical line of the station. Affected by the line quality, load device interference, external radiation interference, etc., the network communication topology is disturbed. The purpose of practice verification in this paper is to use machine learning method to analyze the constantly changing topology information and located interference to eliminate faults.

3. Electromagnetic interference diagnosis treatment scheme

3.1. Overall scheme

Machine learning method is to learn from the data to experience, use the rules of the data to solve the corresponding problems, and use historical data to predict future data trends[5]. The historical node data of the network topology of the courts is collected to form massive topology data, and the law between topology change and electromagnetic interference is found from the network topology, and the key factors affecting the network topology by electromagnetic interference are found to realize the determination of electromagnetic interference. Finally, combined with GPS positioning data and marketing file data, the electromagnetic interference source is located in real time.

The specific design scheme is mainly divided into two parts:(1) by using the network topology data collected by the broadband carrier module to analyze the network topology change law, the scheme identifies the relationship between electromagnetic interference and network topology of courts, and then screens out information of electricity meters with electromagnetic interference.(2) According to the selected meter information and marketing file data, the scheme uses GPS positioning information or geographic location coordinate information to locate the source of electromagnetic interference.

![Electromagnetic Interference Diagnosis Model Solution](image)
3.2. Treatment Procedure

3.2.1. Electromagnetic interference searching. Combining the basic network topology data collected by the broadband carrier module with the frequency data of the interference source, selecting and extracting the electromagnetic interference characteristics, and deeply exchanging the association rules between the electromagnetic interference source and the network topology data of the courts, thereby determining the information of electricity meter that electromagnetic interference occurred.

3.2.2. Electromagnetic interference positioning. Record the electric meter information that has already generated electromagnetic interference, and combine the marketing file data, GPS positioning data, geographical location data, etc. to locate the position where electromagnetic interference occurs in real time, and timely check the interference source to save manpower and material resources and realize the operation of the courts safety management, avoid the loss of electromagnetic interference to the power grid.

4. Establishment of electromagnetic interference diagnosis model

4.1. Electromagnetic interference discovery

4.1.1. Data collection. On the basis of the network topology data of the courts, the data of the electromagnetic interference diagnosis model is collected in two ways. Firstly, the system data of the base area network and the marketing archive data are acquired, and then, basis on the electromagnetic interferon source type, statistics interference source center frequency and meter reading success rate and other data.

4.1.2. Data preprocessing. Data preprocessing techniques can improve the quality of the data, which helps to improve the accuracy and performance of the mining process. Usually, data preprocessing methods are:

- Data normalization
  Normalizing the features of the numeric type can unify all the features into one roughly the same numerical interval. When the model learning speed is the same, the data convergence after normalization is faster than the unnormalized data.

- Data discretization
  When data preprocessing is performed, some feature data needs to be discretized. The usual method is to discretize continuous data by using equal-width interval method and maximum entropy, and divide the value range or value interval of continuous attributes into Between several cells, and assign a discrete symbol to these intervals.

- Data type conversion
  It is necessary to perform type conversion on the collected data, and convert the category type features into numerical features. The one-hot encoding is a conversion that does not exist between the categories, and the ordinal encoding is a conversion that determines the size relationship between the categories. For different feature problems, you can choose the appropriate way to discretize.

4.1.3. Model determination. According to the business scenario of electromagnetic interference, a supervised classification algorithm in machine learning can be considered to solve the problem. Among the classification algorithms, logistic regression(LR), support vector machine(SVM), and XGBoost are usually used, as shown below:

- LR
  LR is a basic binary classification problem. The sigmoid function is used to establish the objective function of logistic regression. Then, the maximum likelihood estimation is used to solve the
problem. By predicting the probability of output and combining the characteristics of sigmoid function, the category judgment is used. In the optimization of loss function, the logarithm method is used to smooth the loss function, and then uses the gradient descent method to continuously adjust the step size and the number of iterations to find the optimal solution of the model[6].

- **SVM**

SVM is a classifier that can simultaneously minimize the empirical error and maximize the geometric edge region. It adds fault tolerance, there will be a certain fault tolerance for the sampled fault, solves the problem of linear inseparability, and then considers the nonlinear factors of the data. In addition, the kernel trick is added, so that the classifier maps the data in the high-dimensional space to the low-dimensional space, and solves the nonlinear separability problem to some extent. SVM establishes the maximum interval of separating hyperplane in the space, so that there is a maximum interval between the point closest to the hyperplane and the hyperplane, and finally the total error of the classifier is minimized[7].

- **XGBoost**

XGBoost uses a 'boosting' idea to integrate learning into multiple weak learners to form a strong learner. The composition of weak learners can be parallel training of the same algorithm, or different algorithms can be used for the same service. The problem is composed. Because the tree structure advantage of the Classification and Regression Tree (CART) is widely used. The CART-based learner is iteratively learned, and the loss function is continuously optimized to minimize the loss function of the classifier[8].

The electromagnetic interference diagnosis model combines various classification algorithms to judge the electromagnetic interference diagnosis model, and comprehensively analyzes the advantages and disadvantages of the electromagnetic interference data characteristics and classification algorithms, and selects the optimal result as the final electromagnetic interference diagnosis model.

4.1.4. **Model training and optimization.** Based on the network topology data of the courts, data collection of the electromagnetic interference diagnosis model is carried out in two aspects. Firstly, the system data of the base area network and the marketing archive data are acquired, and then the data of the network, such as the network access, the self-recovery, and the meter reading rate, can be mined according to the network rules of the network.

Based on the characteristics of electromagnetic interference diagnosis model selection, the training set, verification set and test set are divided into feature datasets, and three kinds of classification algorithms, logistic regression, support vector machine and XGBoost, are used to test and analyze the experimental results. The specific schemes are as follows:

- Scheme 1: Select the LR algorithm to establish an electromagnetic diagnosis interference model based on LR.
- Scheme 2: Select the SVM algorithm to establish an electromagnetic diagnosis interference model based on SVM.
- Scheme 3: Select the XGBoost algorithm to establish an electromagnetic diagnostic interference model based on XGBoost.

Combining the grid search to optimize the parameters, the optimal solution of the three algorithms in the same data set is obtained. The final experimental results are as follows:

| Schemes  | Model Parameter | Training Time  | Training Result |
|----------|-----------------|----------------|-----------------|
|          | Regularization  | Loss Function  | Precision | Recall | F1-Score |
| Scheme 1 | L2              | 0-1 loss       | 2h21min    | 74.11% | 72.52%  | 73.03%  |
| Scheme 2 | L2              | Squared_hinge  | 2h02min    | 88.27% | 85.63%  | 86.93%  |
| Scheme 3 | L1,L2           | Squared_root   | 1h57min    | 93.12% | 96.77%  | 94.91%  |

Table 1. Calculation results of three options.
After comparing and analyzing these three algorithms, the third algorithm can be selected for the final algorithm of the electromagnetic interference diagnosis model. In order to test the generalization ability of the electromagnetic interference diagnosis model, it is considered to re-select the data to verify the model.

4.1.5. Model prediction and verification. Based on the electromagnetic interference diagnosis model, the data of the network topology change in the courts is re-selected for analysis. A total of 30 network topology data of the courts are selected. According to the change of the network topology of the courts, the number of nodes entering the network, the meter reading rate, the frequency and other data, the nodes with electromagnetic interference in the courts are analyzed, and the accuracy of the final model reached 95.38%.

4.2. Electromagnetic interference positioning
According to the model and by using the marketing archive data or GPS positioning data, the model finds the electromagnetic interference source and makes analysis from the following two perspectives:(1) Based on information of the electric meter with electromagnetic interference related to the marketing data file, the model identifies the location of the electricity meter with electromagnetic interference and narrows the scope of electromagnetic interference;(2) By referring to the GPS positioning data, the model realizes real-time positioning of the electricity meter to identify the scope of electromagnetic interference more accurately.

5. Conclusion
Based on the topology data of the carrier network of the station, this paper uses a variety of classification algorithms for analysis, and finally comes up with an electromagnetic interference diagnosis model based on the XGBoost classification algorithm. Firstly, using the network topology information of the courts, the timing sequence topology of the courts is fitted. After analyzing the variation law of the timing sequence topology, the characteristics of the influence of electromagnetic interference on the network topology change are distinguished. The electromagnetic interference is classified by XGBoost to realize the positioning of electromagnetic interference in the courts, and helps field operation and maintenance personnel to quickly deal with faults.

Acknowledgment
Science and Technology Project Support of State Grid co., LTD. No. SGLNJJL00ZHJS1900017

References
[1] Dongjing Bu. (2015) The influence of harmonic pollution on carrier communication and its suppression measures. J. China New Telecommunications., 17(13): 8-10.
[2] Xiyuan Han. The application of power line broadband carrier communication in smart grid. (2012) Yunnan Power Grid Corporation,Yunnan Electrical Engineering Society.
[3] Gangling Tao. (2017) Influence of residential electricity environment on power carrier power margin. Telecom Power Technologies., 34(06): 194-196.
[4] Junchao Zhu, Xiangang Peng, Yong Yang, Zhuangmao Li, Kai Zheng. (2018) A low-voltage power line carrier communication routing method.J. Electron Measurement and Instrument., 55(11): 58-64.
[5] Luwa Hu. (2018) The Quest for Machine Learning. In: Geyue Zhu, People Post Press. PTPRSS.,Beijing.
[6] B.Krishnapuram, L. Carin, M.A.T. Figueiredo, and A.J. Hartemink. (2005) Sparse multinomial logistic regression: Fast algorithms and generalization bounds. IEEE Trans. on Pattern Analysis and Machine Intelligence., 27.
[7] Chih-Wei Hsu and Chih-Jen Lin. (2002) A comparison of methods for multiclass support vector machines. IEEE transactions on Neural Networks 13, pp. 415–425.
[8] Peter Buehlmann and Torsten Hothorn. (2007) Boosting Algorithms: Regularization, Prediction and Model Fitting (with Discussion). Statistical Science., 22(4): 477–505.