Fault Prediction Analysis of Communication Optical Fiber Based on SVM Algorithm

Pengfei Li(✉)

Qinghai Minzu University, Xining, Qinghai, China

Abstract. Optical fiber is the basis of communication network, carrying a huge network traffic, the impact of the cable failure is significant. As a result, the fiber fault prediction is a hot research topic. In this paper, based on the basic parameters and fault information of optical fiber, Support Vector Machine (SVM) model is adopted to classify the faults. Since the cable fault is a small probability event, there is an imbalance in the distribution of the data sample. Use Synthetic Minority Oversampling Technique (SMOTE) algorithm to treat the imbalance data, and then analyze and study by using the SVM classification model.

The results show that the overall classification accuracy of the prediction model is 79.8%, and the prediction sensitivity of the fault fiber is 62.2%. The results show that the model has a good effect, and it can provide a certain auxiliary role for the operation and maintenance of communication fiber.

Keywords: Optical cable fault · SVM algorithm · SMOTE algorithm

1 Introduction

Optical fiber has the advantages of large communication capacity, long distance and strong anti-electromagnetic interference ability, it is widely used in communication network [1]. Multiple optical fibers are combined into optical cables, which are laid in the form of overhead, pipeline, and ground, etc., thus establishing the foundation of the whole communication network. The environment of optical cable laying is complex, once the fault causes interruption, it will have a great impact on the communication network. Therefore, the prediction of optical cable fault is the difficulty and key to the operation and maintenance of communication system, this problem is widely researched [2].

At present, in order to deal with optical cable failure, a series of technologies have been developed, such as optical fiber automatic protection switching

Supported by The research innovation team of the multi-source data fusion and application of Qinghai Minzu University, and Qinghai Minzu University 2018 University-level Science and Technology Research Project under Grant 2018XJQ04.

X. Wang et al. (Eds.): AICON 2021, LNICST 396, pp. 82–86, 2021.
https://doi.org/10.1007/978-3-030-90196-7_8

technology. Optical cable online monitoring system is also widely used, it uses optical time domain reflectometer and other equipment to monitor the running state of optical cable for a long time to realize the functions of optical cable fault warning, analysis, positioning and so on, and provides guarantee for the safe and efficient operation of optical cable network [3,4].

In the main research methods of optical cable fault prediction at present, data mining algorithms are also beginning to be applied, based on the optical power data to carry out the analysis of a large amount for a long time, according to the mathematical characteristics of light power on the time series, optical power time series of high-frequency random factor and trend of low frequency factor analysis using wavelet transform method, optical power trend analysis and forecasting, fiber optical line state warning system [5]. The above methods are expensive to implement and require special hardware and software platforms, which can only be realized by collecting a large amount of data for a long time.

The main contribution of this article is a novel framework for optical cable fault prediction. We propose a method, based on data mining algorithm based on optical fiber cable directly basic parameters and history fault information, with cable length, type, installation, usage and so on as the input features, using SMOTE completion algorithm does not balance data, fault classification training SVM model, prediction results can be a preliminary screening failure probability higher cable, with very low cost for cable operation maintenance efficiency.

The rest of this article is organized as follows. How to select the valuable features form basic data of optical cable is described in Sect. 2. Section 3 solve the problem of data type imbalance. In Sect. 5, we use SVM algorithm to find the optical cable failure, the experimental results are presented. At last, Conclusions are in Sect. 6.

2 Data and Feature Selection of Optical Cable

Based on the basic parameters and historical fault data of all optical cables in the power communication network of a province, this paper carry out the training and verification of the prediction model. There is a total of 1652 optical cables, 121 of them have faults, and the fault samples only account for 7.9% of the normal samples, this is a typical unbalanced data set.

According to the characteristics of the fiber optic cable, delete parts of the features which has nothing to do with the cable fault variables, ultimately determine the input characteristics of the prediction model has: the total number of fiber in the cable, the number of optical fibers already in use, the manufacturer, the fiber length, the laying type, the user, the construction completion time. The model of predicting output is fault state (0 indicates normal, 1 failure).

3 SMOTE Algorithm Processes Unbalanced Data

At present, logistic regression, support vector machine, neural network, K-nearest neighbor, decision tree and other mainstream classification algorithms

in machine learning require the quantity distribution of all categories to be balanced. When the above algorithm is directly applied to unbalanced data samples, the effect is going to be very bad, such as events with low probability such as default, illness, disqualification, failure, etc., which are often classified into most classes, but rare samples with low probability are exactly what the prediction is really concerned about [5–7].

To solve the above problems, it is necessary to add a small number of sample data to improve the model effect. SMOTE is an approach to the construction of classifiers from imbalanced datasets is described. It can add the new sample to the data set artificially according to the Minority samples, which solve the overfitting problem of the model caused by the over-sampling algorithm adopting the strategy of simply copying the samples [8].

The basic principle of SMOTE algorithm: in the Euclidean space of the data point, for each minority sample xi, choose a sample xj randomly from its nearest neighbor, and then choose a randomly selected point on the line between the two points as the newly composed minority sample [9].

In this paper, on the basis of 121 fault optical cable data, 1413 fault data are supplemented by SMOKE algorithm, and the original 1652 optical cable data are expanded to 3065, among which fault and normal account for 50% respectively, providing a balanced data sample for the prediction model.

4 SVM Algorithm

SVM is a class of supervised learning that carries out binary classification of data in the generalized linear classifier, whose decision boundary is the maximum margin hyperplane that can be solved for the learning sample [10]. SVM is first proposed by Cortes and Vapnik in 1995, has shown many unique advantages in solving small sample sizes, nonlinear and high-dimensional pattern recognition and can be generalized to other machine learning problems such as function fitting [11].

SVM method is based on statistical learning theory based on VC dimension theory and structure risk minimum principle, according to the limited sample information in the complexity of the model (i.e. the specific learning Accuracy of training samples, Accuracy) and learning ability to seek the best compromise between, in order to get the best generalization ability [12]. Therefore, SVM is very suitable to solve the problem of optical cable fault prediction in our study.

5 Analysis of Classification Results

Using Scikit-learn machine learning framework [13], we set up a multi-layer perceptron neural network model, the SMOKE algorithm adds balance after 3065 cable data, random selection of 80% of the data samples constitute the training sample, the remaining 20% of the data form the test samples, according to the result of training to determine the suitable network convergence threshold and weights, the SVM model uses the gaussian kernel function, the test data is used to test the model predictions, the effect as shown in Table 1:

Table 1. List of notation

Real value/Predicted	Normal	Fault
Normal	293	5
Fault	119	196

The accuracy rate is a good and intuitive evaluation index to measure the quality of the classification algorithm, but sometimes the high accuracy rate does not mean that the algorithm can solve problems. In this paper, there are only two fault categories of optical cable: 0 means normal, 1 means failure. In the operation and maintenance of optical cable, it is more concerned about whether faults can be found. The omission of faults is more serious than the false detection of normal optical cable. Therefore, the prediction sensitivity of optical cable faults (also known as recall rate [14]) is more important, that is, the proportion of the actual fault samples predicted as faults. According to Table 1, the prediction classification results of the cable fault warning model are calculated as follows:

$$Precision = (293 + 196) / (293 + 5 + 119 + 196) = 79.8\%$$
$$Recall = 196 / (196 + 119) = 62.2\%$$

The classification effect of this model has an overall accuracy rate of 79.8%, and the recall of the fault optical cable reaches 62.2%, with relatively high accuracy. Obviously, the optical cable fault probability model established in this study can provide a certain auxiliary role for the operation and maintenance of optical cable.

6 Conclusion

In this paper, the optical cable fault warning model is proposed, and the operation and fault data of the optical cable in a provincial power communication network are selected for example analysis. The SVM algorithm is used to predict the fault based on the key operation indexes such as the length, type and laying mode of the optical cable. By analyzing, it is found that the traditional classification algorithm has a poor classification effect on this kind of unbalanced data. Therefore, SMOTE algorithm is selected for the balance treatment of the sample test data, and the predicted effect after treatment is better. In this study, under the condition of not increasing the cost of hardware and software, the cable fault warning model is used to predict, and the cable fault possibility is preliminarily screened to strengthen the maintenance in advance to reduce the fault, promoting the cable operation and maintenance work from the empirical type to the analysis type, engaged in the passive type to advance the active type transformation.

In the following work, more variable characteristics can be added to the cable fault warning model for research, such as optical receiving power, optical

transmission power, machine room operating parameters, cable off-line type, etc., to further improve the prediction sensitivity of the model, better guide the cable operation and maintenance work.

References

1. Keiser, G.: Optical Fiber Communications. Wiley Encyclopedia of Telecommunications (2003)
2. Wang, Z., Zhang, M., Wang, D., et al.: Failure prediction using machine learning and time series in optical network. Opt. Exp. **25**(16), 18553–18565 (2017)
3. Shaneman, K,, Gray, S.: Optical network security: technical analysis of fiber tapping mechanisms and methods for detection & prevention. In: IEEE MILCOM 2004. Military Communications Conference, 2004, vol. 2, pp. 711-716. IEEE (2004)
4. Zhang, H., Dai, G.L.: Design of optical fiber communication network monitoring and detection system based on address resolution protocol cheats. Optik **127**(23), 11242–11249 (2016)
5. Tao, Q., Wu, G.W., Wang, F.Y., et al.: Posterior probability support vector machines for unbalanced data[J]. IEEE Transactions on Neural Networks **16**(6), 1561–1573 (2005)
6. Searle, S R.: Linear Models for Unbalanced Data. John Wiley & Sons, New York (2006)
7. Mani, I., Zhang, I.: kNN approach to unbalanced data distributions: a case study involving information extraction. In: Proceedings of Workshop on Learning from Imbalanced Datasets, vol. 126. United States: ICML (2003)
8. Chawla, N.V., Bowyer, K.W., Hall, L.O., et al.: SMOTE: synthetic minority oversampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
9. Fernández, A., Garcia, S., Herrera, F., et al. : SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. J. Artif. Intell. Res. **61**, 863–905 (2018)
10. Joachims, T.: Making large-scale SVM learning practical[R]. Technical report (1998)
11. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local SVM approach. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004 (ICPR 2004), vol. 3, pp. 32–36. IEEE (2004)
12. Leslie, C., Eskin, E., Noble, W.S.: The spectrum kernel: a string kernel for SVM protein classification. Biocomputing **2001**, 564–575 (2002)
13. Pedregosa, F., Varoquaux, G., Gramfort, A., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
14. Davis, J., Goadrich, M.: The relationship between Precision-Recall and ROC curves. In: Proceedings of the 23rd International Conference on Machine Learning, pp. 233–240 (2006)