Journal of Electrotechnology, Electrical Engineering and Management (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jeeem.2024.070120
ISSN 2560-6697 Vol. 7 Num. 1

Research on Distribution Engineering Information Recognition Technology Based on Improved RF Algorithm

Ying Ding[1,a,*]

[1]IT Solution Consulting, MOYI Inc, New York, NY, USA
[a]hanyue0409@yahoo.com
[*]Corresponding author

Keywords: Random Forest, Engineering Information, Kernel Principal Component Analysis, Precise Weighting, Classification Recognition, Low-Voltage Distribution Network

Abstract: With the deep integration of the Internet of Things, big data, and artificial intelligence technology, smart grids are moving towards higher levels of automation and intelligent management. In today's rapidly changing smart grid technology, improving the intelligence and accuracy of distribution network management has become an urgent need for industry development. In response to this challenge, this study proposes a low-voltage distribution network engineering information recognition technology based on an improved random forest algorithm, aiming to solve the problem of discrepancies between the settlement engineering quantity and the actual engineering quantity in the settlement management of the power system distribution network. Extract data features through kernel principal component analysis algorithm, construct feature vectors of engineering information, and optimize random forest parameter selection through precise weighted decision tree and particle swarm optimization algorithm. The experimental results show that the algorithm can effectively identify abnormal engineering information, with higher accuracy and efficiency, providing new ideas and practical experience for the development of intelligent distribution network management technology.

1. Introduction

In recent years, the demand for power grid construction in China has continued to rise. However, what does not match this growth trend is the completion settlement management status of power infrastructure construction projects. Especially in the field of low-voltage distribution networks, due to the large number of power equipment, the amount of data information generated increases the burden of settlement management. In the current settlement management process, the construction unit is responsible for preparing settlement documents, while the owner's project department is responsible for reviewing the work. However, a series of problems have been exposed in practical work: the lack of high-level distribution network technical personnel in the construction unit and the owner's project department has threatened the accuracy and consistency of settlement documents;

The complexity of low-voltage distribution network construction projects leads to a wide range of settlement documents, and the measurement process is prone to errors and abnormal data caused by inconsistent accuracy; However, the current data collection, verification, statistical calculation, and progress management rely on manual labor, which is inefficient and difficult to adapt to the long-term and short-term development planning needs of power grid construction.

In response to these challenges, this study focuses on improving the random forest algorithm, aiming to achieve accurate identification of low-voltage distribution network engineering information, thereby reducing settlement errors and improving construction quality. Through feature extraction and recognition, abnormal data in engineering information can be effectively identified, avoiding the inefficiency and error problems caused by traditional manual methods. This study not only focuses on solving the pain points of current power grid settlement management, but also opens up an innovative path for the progress of intelligent distribution grid management. Through continuous innovation and optimization of technological means, it is expected to bring more efficient and accurate solutions to the field of power system distribution network management, and contribute to the sustainable development of China's power grid construction.

2. Related Research

C Liu et al. proposed a dynamic prediction algorithm for low-voltage distribution network line losses in the article[1], which combines classification decision trees with marketing data. The experimental results show that compared with the three methods, the data loss rate of this algorithm is less than 0.2, the root mean square relative error is less than 0.02, and the average fitting degree is higher than 0.08. The results have proven that future predictions will be implemented in smart cities. Z Tang proposed a short-term line loss prediction algorithm for low-voltage distribution networks based on Kmeans LightGBM[2].A low-voltage distribution network operation data quality evaluation system was established based on the Hadoop platform. The feature dimension was extended through feature engineering, and feature dimensions that do not have multicollinearity and are highly correlated with line losses were selected. Data normalization was performed again. After comparison, the prediction accuracy of the model was higher than BPNN, FOA-SVR, and traditional LightGBM.

Laxmipriya Samal et al.[3]combined the K-means prior feature selection (KAFS) algorithm with statistical techniques to propose a new feature vector for classifying three-phase power quality disturbance (PQD) events. The ability of Short Time Fourier Transform (STFT) as a time-frequency tool was evaluated using the KAFS algorithm for the described task. Choose a Random Forest (RF) classifier to validate the effectiveness of the proposed feature vectors. From our results, it can be seen that the use of KAFS's new optimized feature vectors has indeed improved recognition accuracy.

3. Information Recognition Model for Pressure Distribution Power Grid Engineering

3.1 Random Forest Algorithm

The Random Forest (RF) algorithm is an extension of the Bootstrap Aggregation (Bagging) algorithm. This algorithm has two functions: selecting the trained samples with dropout and selecting samples without dropout for different attributes, effectively improving global search performance and achieving high classification accuracy. The weak classifier used in the RF algorithm is the decision tree, and each decision tree uses the Bagging algorithm to sample the training samples. At the same time, it can also use the random subspace algorithm to sample the attributes. In addition, since any decision tree in RF is independent and distinct, it has strong adaptability to training samples.

The random forest (RF) classifier is chosen to validate the efficacy of the proposed feature vectors. The novel optimised feature vectors using the KAFS have indeed enhanced the recognition accuracy as revealed from our results[4].

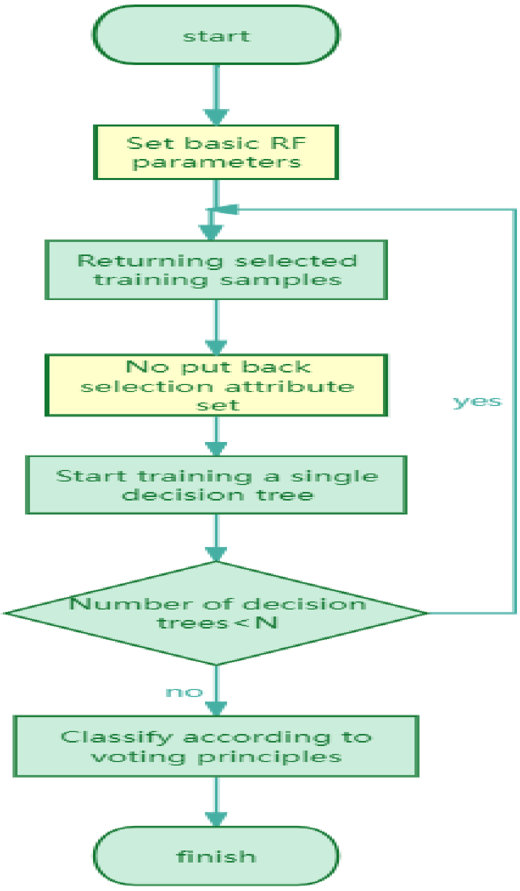

Figure 1: RF Algorithm Process

The core process of in-depth research on the RF algorithm is shown in Figure 1: ① Set the initial parameters of the RF algorithm, including the number of samples N and the number of attributes I, and determine the pruning threshold of the decision tree during the training process. ② By selecting samples from the training set with put back sampling and selecting attributes without put back sampling, sufficient data support is provided for the training of a single decision tree Input the selected samples into a single decision tree for training, and prune the decision tree according to the preset pruning threshold. If the number of trained decision trees is lower than the set value of N, the above steps are iterated until the expected number is reached. In this study, the classification results of multiple decision trees are integrated by voting principle, and this pattern is selected as the final classification result of RF.

The optimization of this process is aimed at improving the accuracy and efficiency of information recognition in power distribution engineering. In practical applications, parameters should be adjusted according to specific situations to ensure that the model can better adapt to different distribution engineering information scenarios. Meanwhile, the reasonable selection of training samples and accurate extraction of attributes are also key steps that can effectively improve the performance and generalization ability of the model. Through these optimization measures, a more reliable and efficient distribution engineering information recognition model can be constructed, providing stronger support for power grid management.

3.2 Improvement of RF Voting Principle under Precise Weighted Optimization

In traditional RF algorithms, each decision tree has the same voting weight, and this simple voting principle may not fully utilize the classification ability of each tree, especially in the presence of performance differences. When the performance of decision trees is different, traditional methods may lead to poor performance decision trees having a significant impact on the final result, thereby reducing the overall classification accuracy. Furthermore, when multiple decision trees generate the same number of votes, traditional methods cannot effectively solve the decision-making problem.

To address these issues, a precise weighted RF voting principle improvement method is proposed. By determining weights based on the classification accuracy and other evaluation indicators of each decision tree on the validation set, the classification performance of each tree can be more accurately measured, and weighted voting can be conducted accordingly. In this way, decision trees with better performance will be given greater weight, thus playing a greater role in the final classification results.

This precise weighting method can effectively improve the classification accuracy of the RF algorithm, while also enhancing the robustness and stability of the model. By fully considering the contribution of each decision tree, the information of the model can be better utilized, thereby improving the credibility of the classification results. This is of great significance for the development of information recognition technology in distribution engineering and the intelligence of power grid management.

4. Experimental Analysis

4.1 Experimental Environment and Data Source Analysis

Table 1: Dataset Information

Sample information	Number of training sets/piece	Number of test sets/piece	Number/type of attributes	Number/type of categories
Project operation information	1000	726	10	3
Construction supervision information	1000	542	8	5
Construction progress information	1000	482	6	2
Contract information	1000	816	12	4
Safetyre gulatory information	1000	138	6	10
Materialand equipment information	1000	677	18	32
Presettlement preparation information	1000	619	8	4

The experimental data covers multiple aspects of low-voltage distribution network engineering, including project operation, construction supervision, safety management, etc., as shown in Table 1. These pieces of information are crucial for studying and improving the distribution engineering information recognition technology using RF algorithms. By comprehensively analyzing these data, we can better understand the operational status of power distribution engineering projects and identify opportunities for improvement. Therefore, fully utilizing these information resources will help optimize algorithms to improve the accuracy and efficiency of information recognition.

The experimental part mainly improves the random forest algorithm through programming. As shown in Figure 2.

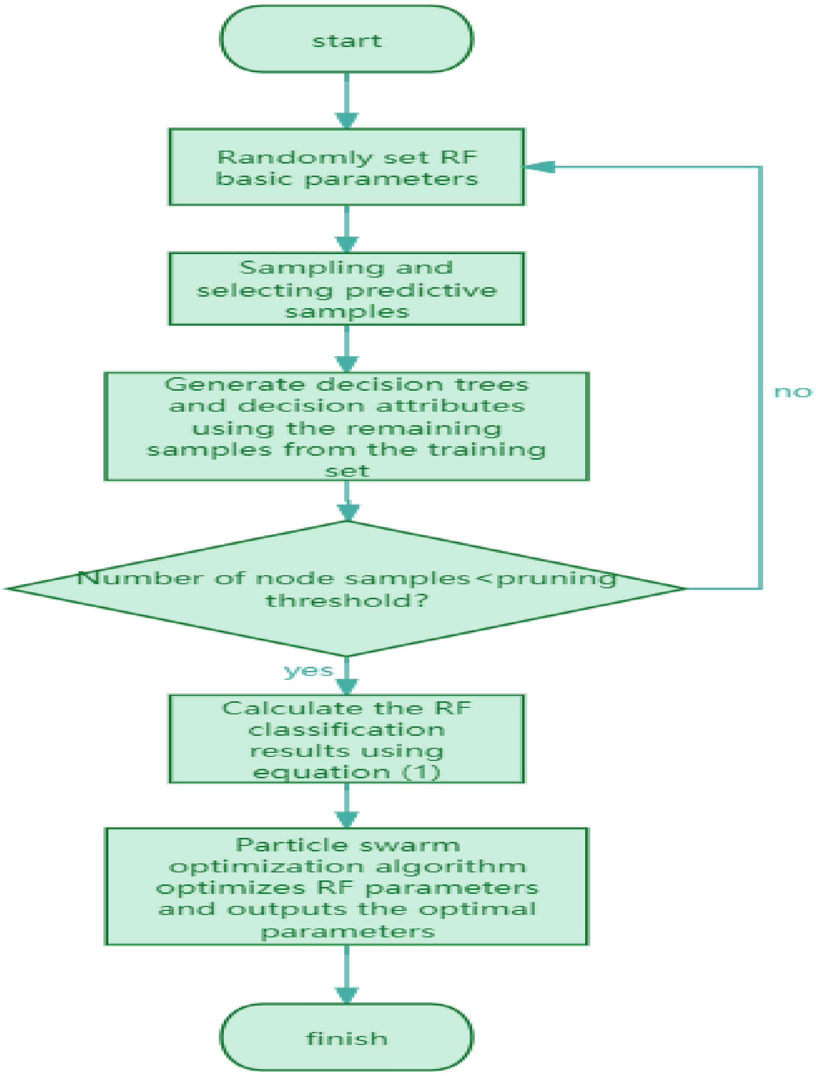

Figure 2: Particle Swarm Optimization RF Process

4.2 Analysis of Information Identification Results for Low Voltage Distribution Network Engineering

The improved RF algorithm was applied to the identification of engineering information in low-voltage distribution networks, achieving accurate identification of abnormal information. Meanwhile, the algorithm was compared and analyzed with BP neural network, SVM, RBF neural network, and PSO optimized algorithm. The recognition results obtained are shown in Table 2.

These experimental results demonstrate significant performance advantages of the improved RF algorithm in low-voltage distribution network engineering information recognition. By comparing with other commonly used algorithms, the effectiveness of the improved RF algorithm in identifying abnormal information was verified. These results not only demonstrate the effectiveness of the algorithm, but also provide useful reference and inspiration for the research of information recognition technology in distribution engineering.

Table 2: Identification Results

algorithm	Training set recognition accuracy(%)	Test set recognition accuracy(%)
BP	81.2	75.5
PSO-BP	87.3	82.5
SVM	88.5	83.5
PSO-SVM	90.2	88.0
RBF	83.9	79.0
PSO-RBF	86.2	81.5
RF	90.2	86.5
PSO-RF	93.8	91.0

5. Conclusion

This article proposes an improved random forest algorithm to address the complexity of information management in the construction process of low-voltage distribution network engineering, in order to address the challenge of information recognition. By utilizing the KPCA algorithm to extract features of engineering information, and combining precise weighted improved voting principles and particle swarm optimization algorithm to optimize the parameters of the random forest model, an effective low-voltage distribution network engineering information recognition technology scheme has been successfully designed. The experimental results show that the proposed method is significantly superior to other methods in terms of recognition accuracy, and can accurately identify engineering information of low-voltage distribution networks, which has important engineering practical value.

References

[1] Liu C, Fu L, Li H, et al.Dynamic Prediction Algorithm for Low-Voltage Distribution Network Power Loss in a Smart City Based on Classification Decision Tree and Marketing Data[J].Journal of Testing and Evaluation: A Multidisciplinary Forum for Applied Sciences and Engineering, 2023.

[2] Tang Z , Xiao Y , Jiao Y ,et al.Research on Short-Term Low-Voltage Distribution Network Line Loss Prediction Based on Kmeans-LightGBM[J].Journal of circuits, systems and computers, 2022.DOI:10.1142/S0218126622502280.

[3] Samal L , Palo H K , Sahu B N .The recognition of 3-phase power quality events using optimal feature selection and random forest classifier[J].International journal of computational vision and robotics, 2023.

[4] Lai X , Cao M , Liu S ,et al.Low-voltage distribution network topology identification method based on characteristic current[C]//2021 6th Asia Conference on Power and Electrical Engineering (ACPEE).2021.DOI: 10.1109/ ACPEE51499.2021.9437092.