ISSN 2616-5775 Vol. 7, Issue 12: 9-15, DOI: 10.25236/AJCIS.2024.071202

Decision Tree Algorithm Based Tidal Flow Calculation for Distribution Networks

Tianxiang Huang*

School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan, 243032, China
**Corresponding author: 18755501126@163.com*

Abstract: *As an important part of the power system, the accuracy of the current calculation of the distribution network is of great significance for the operation stability analysis and optimization design of the power system. Traditional current calculation methods include Newton-Raphson method, PQ decomposition method, etc. Due to the difficulty in obtaining line parameters and complex source-load characteristics of distribution networks, it is difficult to meet the demand for accurate current calculation by applying the traditional current calculation methods directly. Aiming at the problems that the traditional trend calculation methods are not applicable to the radial structure, open-loop operation, and the large number of nodes and branches of distribution networks, this article proposes a decision tree algorithm based distribution network trend calculation method to realize the accurate trend analysis of distribution networks from a data-driven perspective. First, a dataset of distribution network trend data is obtained; then, based on the collected data, a decision tree model is trained to realize accurate distribution network trend analysis without the need of distribution network line parameters; finally, the trained model is applied to the IEEE33 node distribution network system, and the validity and accuracy of the model are verified by predicting the trend of this system.*

Keywords: *Distribution networks; Current calculation; Decision tree modeling; Data-driven; IEEE33 node system*

1. Introduction

Distribution network trending calculations are used to determine node voltages and branch power flows, which are usually solved by iterative methods due to nonlinear relationships. In the case of complex structures, large load variations, and many distributed power supply accesses, the calculation may not converge, which is affected by the initial values, parameter settings, and network topology [1]. The intermittency and complex topology of distributed power sources increase the computational difficulty and uncertainty. The commonly used Newton-Raphson method converges quickly but is sensitive to the initial value and has high computational cost. Various improved algorithms have been proposed to address these issues.

With the rapid development of smart grid and other related technologies, the methods of power system distribution network trend calculation have made great progress [2]. A linear optimal current model based on the data-driven linear current model is proposed in the data-driven optimal current model for distribution networks under partially observable conditions. The idea of conduction matrix transformation in WARD equivalence theory is borrowed for system topology equivalence. The unobservable nodes are considered as external nodes and the post-equivalent network topology is obtained without relying on any line parameters [3]. In the iterative implicit linearization based mesh distribution network current calculation and its optimal current model, the iterative implicit linearization current model and its optimal current model are constructed. The model considers the nonlinear trend manifold M(Manifold) as an implicit algebraic relationship between the node voltages and the node injected power, followed by a local approximation using tangent plane pairs and iteratively updating the linearization points to improve the approximation accuracy of the linear model [4]. A data-driven novel power system trend analysis method is designed in the data-driven novel power system trend analysis method, which considers the node load and new energy output fluctuation of the novel power system as feature inputs, makes full use of the existing massive data, and applies data-driven techniques for big data mining [5]. Literature [6] proposed a linearized tidal current calculation method for distribution networks considering load voltage static characteristics and PV nodes, which has a concise expression

ISSN 2616-5775 Vol. 7, Issue 12: 9-15, DOI: 10.25236/AJCIS.2024.071202

and can directly find the voltage distribution [6]. Literature [7] improved the BPNN model by combining three characteristic variables of the power system for training, optimized the network using GA and ADAM algorithms, effectively improved the efficiency and accuracy of the tidal current calculation, and solved the problem of large-scale data processing [7].

To address the above problems, this paper first introduces the traditional trend calculation methods and analyzes the limitations of the traditional trend calculation methods; subsequently, a decision tree-based distribution network trend calculation model is proposed, which learns the nonlinear mapping relationship between the load information of each node of the distribution network and the voltage of each node as well as the total network loss by training the data-driven model, and then realizes the accurate sensing of distribution network voltage and network loss without relying on the line parameters of the distribution network.

2. Traditional distribution network/power system trend calculation

The traditional power system trend calculation is a basic electrical calculation to study the steady state operation of the power system, and its task is to determine the operation state of the whole system according to the given operation conditions and network structure, such as the voltage (magnitude and phase angle) on each bus, the power distribution in the network, and the power loss, etc. The traditional power system trend calculation is a basic electrical calculation to study the steady state operation of the power system. The traditional power system current calculation is usually solved by an iterative method, and the commonly used iterative methods include the Newton-Raphson method and the P-Q decomposition method.

2.1 Newton-Raphson method

In complex distribution network trend calculation, for a power network with n independent nodes, Y_{ij} is the node conductance matrix element . In Cartesian coordinates, the node voltage is:

$$\dot{U}_i = e_i + jf_i \tag{1}$$

The node guide elements are:

$$Y_{ij} + G_{ij} + jB_{ij} \tag{2}$$

The node voltage and node conductance equations are brought into the power equations for active and reactive power, which are further fine-tuned to obtain the corresponding expressions for the nonlinear equations of the power network:

$$\begin{cases} \Delta P_i = P_i - \sum_{j=1}^{j=n} \left[e_i \left(G_{ij} e_j - B_{ij} f_j \right) + f_i (G_{ij} f_j + B_{ij} e_j) \right] \\ \Delta Q_i = Q_i - \sum_{j=1}^{j=n} \left[f_i \left(G_{ij} e_j - B_{ij} f_j \right) - e_i (G_{ij} f_j + B_{ij} e_j) \right] \\ \Delta U^2 = U_i - (e_i^2 + f_i^2) \end{cases} \tag{3}$$

Based on the computational idea of the Newton-Raphson method, the modified equation based on the Newton-Raphson method can be obtained by stepwise linearization of Eq. (3) using first-order Taylor:

$$\begin{bmatrix} \Delta P \\ \Delta Q \\ \Delta U^2 \end{bmatrix} = \begin{bmatrix} H & N \\ J & L \\ R & S \end{bmatrix} \begin{bmatrix} \Delta f \\ \Delta e \end{bmatrix} \tag{4}$$

where the Jacobi matrix coefficient matrix elements are respectively:

ISSN 2616-5775 Vol. 7, Issue 12: 9-15, DOI: 10.25236/AJCIS.2024.071202

$$H_{ij}=\frac{\delta\Delta P_i}{\delta f_j}\quad N_{ij}=\frac{\delta\Delta P_i}{\delta e_j}\quad J_{ij}=\frac{\delta\Delta Q_i}{\delta f_j} \tag{5}$$

$$L_{ij}=\frac{\delta\Delta Q_i}{\delta e_j}\quad R_{ij}=\frac{\delta\Delta U_i^2}{\delta f_j}\quad S_{ij}=\frac{\delta U_i^2}{\delta e_j} \tag{6}$$

According to the power network structure and network parameters, the node conductance matrix is formed, given the initial value of each node voltage $e_i^{(0)}$, $f_i^{(0)}$, the initial value of each node voltage is substituted into Eq. (3), and the offsets of each node power and node voltage in Eq. (4) are found out and looped and iterated continuously, and each iteration updates each node's voltage value and power value until convergence or predefined error accuracy is reached [8].

2.2 P-Q decomposition method

The P-Q decomposition method is simplified on the basis of the Newton-Raphson method, and the power equation is decomposed into the active power equation and the reactive power equation to be solved separately, which significantly improves the computational speed. The P-Q decomposition method takes advantage of the characteristics of the AC high-voltage grid in which the line reactance is much larger than the resistance, and the transmission of active power is mainly affected by the node voltage phase, and the transmission of reactive power is mainly affected by the node voltage amplitude, the law of the node voltage amplitude, and the method is simplified by the modified Newton-Raphson method. The correction equation of Newton-Raphson method is simplified. The corrected equation of the Newton-Raphson method is rewritten to the polar coordinate case, and the corresponding equation is:

$$\begin{bmatrix}\Delta P\\ \Delta Q\end{bmatrix}=\begin{bmatrix}H & N\\ J & L\end{bmatrix}\begin{bmatrix}\Delta\delta\\ \Delta V/V\end{bmatrix} \tag{7}$$

Since the value of reactance in a power network is usually much larger than the value of resistance, the nodal active power depends on the magnitude of the nodal voltage magnitude and the reactive power depends on the phase angle variation of the nodal voltage. Equation (7) can be further simplified as:

$$\begin{aligned}\Delta P&=H\cdot\Delta\delta\\ \Delta Q&=L\cdot\Delta V/V\end{aligned} \tag{8}$$

From Eq. (8), the simplified equation is a linear equation of order $(n-1)$, replacing the linear equation of order $2(n-1)$ in Eq. (7), which reduces the workload of iterative calculations by separating the active power P and the reactive power Q [8].

2.3 Shortcomings of the traditional method of solving tidal current calculation equations

The convergence performance of the Newton-Raphson algorithm relies heavily on the choice of initial iteration points. If the initial iteration point is not properly selected, it may lead to slow convergence or even non-convergence of the algorithm. In practice, choosing the appropriate initial iteration point may require additional computation and analysis, increasing the complexity and computational cost of the algorithm. During the iteration process, the Newton-Raphson algorithm needs to constantly recalculate the Jacobi matrix (for power system tidal current calculations) due to the fact that the Jacobi matrix changes during the iteration process. The computation of the Jacobi matrix involves a large number of derivation operations and the matrix itself may be asymmetric, which further increases the computational effort. The large amount of computational work not only increases the time complexity of the algorithm, but may also affect the real-time performance of the algorithm. This may result in the algorithm not being able to meet the practical requirements in situations where fast response is required, such as power systems.

The convergence and computational accuracy of the P-Q decomposition method, which has been simplified on the basis of the Newton-Raphson method, are greatly affected by the system parameters. When the system parameters do not satisfy the assumptions of the P-Q decomposition method (e.g., the

ISSN 2616-5775 Vol. 7, Issue 12: 9-15, DOI: 10.25236/AJCIS.2024.071202

phase angle difference between the voltages at the two ends of the line is large, and the conductance element corresponding to the reactive power of the node does not differ much from the self-conductance of the node, etc.), the convergence speed and computational accuracy of the algorithm may be affected. In addition, the P-Q decomposition method adopts a linear convergence property, which is relatively slow compared with the square convergence speed of the Newton-Raphson algorithm. Especially at the beginning of the iteration, the convergence speed may be even slower. In this paper, a decision tree algorithm based distribution network trend calculation method is proposed to effectively solve the above problems.

3. Decision Tree Algorithm Based Tidal Flow Calculation for Distribution Networks

Decision tree modeling is a powerful machine learning algorithm widely used in tasks such as classification, regression, and feature selection. It is based on a tree structure, and it constructs a model that predicts the value of a target variable by recursively selecting the optimal features for dataset segmentation.

A decision tree model consists of nodes and edges, where nodes are categorized into internal nodes and leaf nodes. Internal nodes represent a test or decision point for a feature attribute, while leaf nodes represent the final decision result (for classification problems) or predicted value (for regression problems). Edges, on the other hand, connect nodes and represent the partitioning of the dataset into different subsets based on different values of the feature attributes.

The central step in building a decision tree model is to determine the optimal split attributes. As the segmentation continues, the samples contained in the branch nodes of the decision tree should belong to the same category as far as possible, which means that each time an attribute is selected for segmentation, the 'purity' of the subsequent nodes should be maximized. In order to measure the purity of the set of samples, the index 'information entropy' is usually used, the lower the value of information entropy, the higher the purity of the set of samples. Information entropy is defined as:

$$E(D) = -\sum_{k=1}^{|y|} p_k \log_2 p_k \tag{9}$$

where D is the set of samples at a node in the process of constructing a task decision model, p_k is the proportion of the kth class of samples in that set of samples, and y is the number of classes of samples [9].

Assuming that there is a discrete attribute a (e.g., task level) in the influence parameter of the task decision model, and that a has V possible values (e.g., general, urgent, and very urgent), node partitioning of the sample set using this attribute will result in V branching nodes. Assigning weights to the branch nodes, the more the number of samples the more the influence of the nodes. The information gain obtained by dividing the sample set D using attribute a is:

$$G(D,a) = E(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|} E(D^v) \tag{10}$$

The greater the information gain corresponding to attribute a, the higher the 'purity improvement'obtained by using this attribute to classify the sample set.

The decision tree is generated in order to reduce the complexity of the decision tree, the information splitting term $\psi(X,s)$ is used to guide the generation of the decision tree, which is calculated as:

$$\psi(X,s) = -\sum_{i=1}^{k} (\frac{X}{s})_i \tag{11}$$

where X is the dataset that serves as the training sample and an attribute in X is denoted by s.

According to equation (11), $\psi(X,s)$ is the entropy of X with respect to attribute s. The

ISSN 2616-5775 Vol. 7, Issue 12: 9-15, DOI: 10.25236/AJCIS.2024.071202

smaller its value indicates that the distribution uniformity of X on s is worse, so the distribution is adjusted by this value as a way to ensure the generation effect of the decision tree.

Based on Eq. (11), the information gain ratio is calculated to obtain the classification decision tree generation result, which is given by the formula:

$$\eta(X,s)=\frac{\xi(X,s)}{\psi(X,s)} \tag{12}$$

According to equation (12), the highest gain ratio attribute is obtained and defined as the splitting attribute, after determining this attribute, the optimal splitting point is determined to generate the branches of the decision tree, and the above steps are performed cyclically, i.e., the decision tree generation is completed [10].

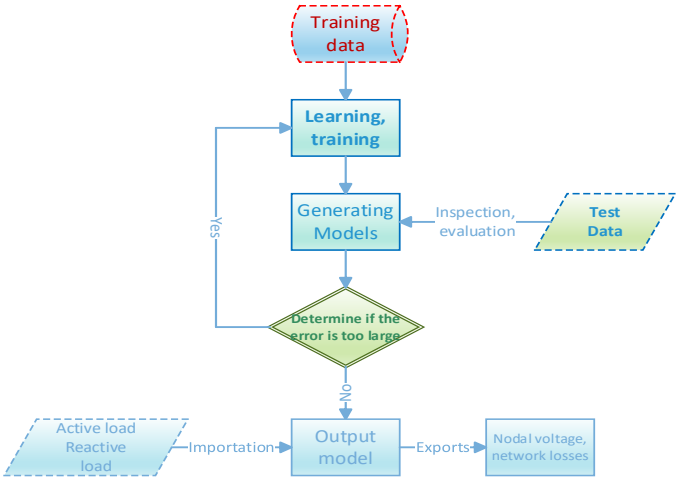

Figure 1: Model training flowchart

Load data, voltage data and network loss data of the distribution network under different operating conditions are collected. Then, a decision tree algorithm is used to train these data to identify and learn the intrinsic connection between node loads and voltage and network losses. During the training process, the decision tree model constructs a series of decision rules by dividing the data set recursively, and these rules can effectively capture the impact of load changes on voltage and network losses. Figure 1 is the flow chart of model training.

4. Calculus analysis

The arithmetic examples in this paper use the IEEE33 system.The network structure of the IEEE33 node distribution system is shown in the Fig.2 in a radial shape with 33 nodes and 32 branches connecting them. Where the head node is viewed as the balancing node. The reference power in the simulation is taken as 10MVA, the reference voltage is taken as 12.66kV and the convergence accuracy is set to 1×10^{-6}.

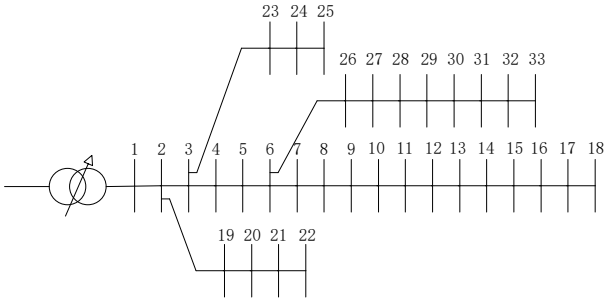

Figure 2: IEEE33 Node Distribution System Network Architecture

ISSN 2616-5775 Vol. 7, Issue 12: 9-15, DOI: 10.25236/AJCIS.2024.071202

To illustrate the approximation accuracy, the Root Mean Squared Error (RMSE) is introduced here, which is calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{13}$$

where n is the number of samples, y_i is the true value of the ith sample, and $\hat{y}_i$ is the predicted value of the ith sample.

In addition, Mean Absolute Error (MAE) was introduced, which is calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{14}$$

The RMSE and MAE of the predicted values of model node voltages and network losses are shown in Table 1:

Table 1: RMSE and MAE of model node voltage and network loss predictions

	RMSE	MAE
Nodal Voltage	0.0013	0.0308
Net Loss	0.0059	0.0647

Figure 3 shows a comparison of predicted and actual node voltage:

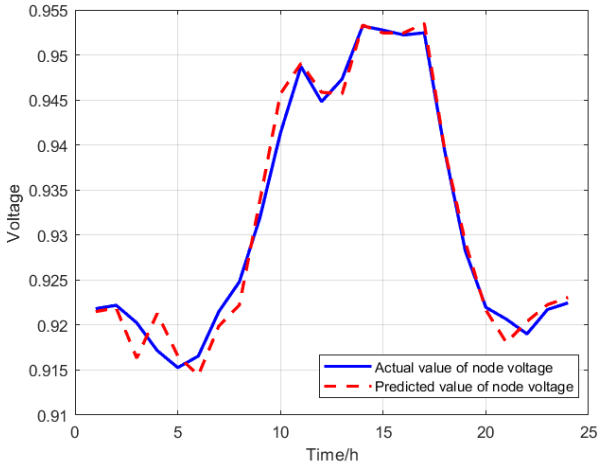

Figure 3: Comparison of predicted and actual node voltage

Figure 4 shows a comparison of predicted and actual network losses:

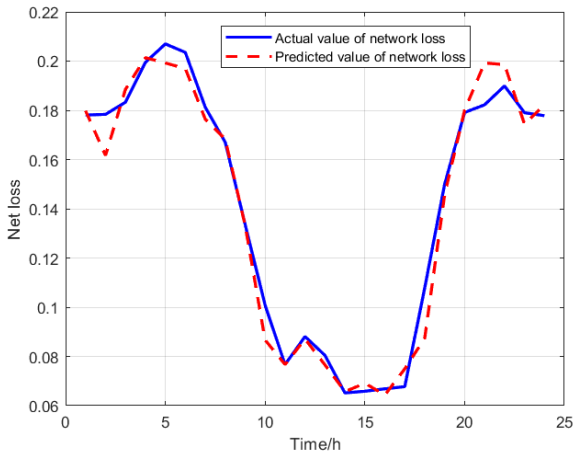

Figure 4: Comparison of predicted and actual network losses

ISSN 2616-5775 Vol. 7, Issue 12: 9-15, DOI: 10.25236/AJCIS.2024.071202

5. Conclusions

According to the decision tree algorithm-based distribution network trend calculation method proposed in this paper, the effectiveness and accuracy of the method are verified through the trend prediction experiments on the IEEE33-node distribution network system. Compared with the traditional tidal current calculation method, the proposed method is able to realize the accurate tidal current analysis under the complex source-load characteristics of distribution networks without the need to accurately obtain the distribution network line parameters. Experimental results show that the data-driven decision tree-based model can maintain high computational accuracy despite the large number of processing nodes and branches. Therefore, the method proposed in this paper provides an effective alternative for tidal current calculation in distribution networks.

References

[1] Wu Xiaohui, Wang Zhongwen, Jiang Jiahe, et al. Research on the impact of distributed power supply on distribution network trend calculation [J]. Science and Technology Information, 2024, 22(13): 84-86.
[2] Zhang Wei, Wang Bin, Wang Xianli, et al. Non-iterative distribution network trend calculation based on BPNN [J/OL]. China Electric Power, 1-8[2024-08-14].
[3] Li Penghua, Song Zhuoran, Wu Wenchuan. Data-driven optimal tidal current model for distribution networks under partially observable conditions [J]. China Electric Power, 2023, 56(12): 51-57.
[4] Zhao Fei, Fan Xuejun, Li Yalou, et al. Iterative implicit linearization based mesh distribution network current calculation and its optimal current model [J]. Chinese Journal of Electrical Engineering, 2024, 44(06): 2197-2208.
[5] Long Yun, Lu Youfei, Zhao Hongwei, et al. A novel data-driven power system trend analysis method [J]. Intelligent Power, 2023, 51(09): 22-30.
[6] Zhao Enlai, Hu Yu, Guo Jun, et al. Linearized tidal current calculation of distribution networks considering load-voltage static characteristics and PV nodes [J]. Electrical Drives, 2024, 54(05): 56-63.
[7] Liu Huijia, Feng Ling, Ai Can. BPNN distribution network trend calculation based on GA-ADAM optimization [J]. Electronic Measurement Technology, 2023, 46(24): 84-92.
[8] Huang Rong. Joint iterative strategy for distribution network trend calculation based on Newton-Raphson method and P-Q decomposition [J]. Automation Application, 2023, 64(22): 55-57.
[9] Liu YZ. Autonomous Task Decision Making Method for Robot Ecosystem Based on Decision Tree Algorithm [C]// Tianjin Electronic Industry Association. Proceedings of the 2024 Annual Conference of Tianjin Electronics Industry Association. Tianjin Seven One Two Communication and Broadcasting Co. Ltd, 2024: 8.
[10] Sheng Lijun, Chen Shiqi. Application of decision tree algorithm in the redundancy of ship autonomous cruise data[J]. Ship Science and Technology, 2024, 46(12): 157-161.