A three-stage random forest integrating fuzzy matrix method in low frequency oscillation classification of power system

Miao Yu[a,b,c], Shuwei Yang[a,b], Fang Shi[d], Sizhuo Gong[a,b], Di Yang[a,b], Zihao Lin[a,b]

a. *School of Mechanical-electronic and Vehicle Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China*

b. *Beijing Engineering Research Center for Building Safety Inspection, Beijing 100044, China*

c. *State Key Laboratory of Power System Operation and Control, Department of Electrical Engineering, Tsinghua University, 100084, China*

d. *School of Electrical Engineering, Shandong University, Jinan, 250061*

Abstract: In view of the problems that the early warning process of low frequency oscillation in power system is vulnerable to the influence of complex grid environment. The classification speed is slow due to the large amount of processing data in the classification process. A three-stage random forest based on a fuzzy matrix method is proposed in this paper to improve the accuracy and the classification speed of low frequency oscillation early warning in power system. Firstly, the fuzzy matrix comprehensive evaluation is carried out by PMU data, and the evaluation score S will be obtained to determine whether low-frequency oscillation occurs and makes a quick warning. Then, the data is processed by Synchronous Wavelet Transform (SWT), and the damping ratio and attenuation factor of the data are obtained. Furthermore, Random Forest 2(RF 2) and RF 3 are used to judge the type of low frequency oscillation. Finally, simulation results show that the comprehensive fuzzy matrix improves the accuracy of low-frequency oscillation early warning, and the three-stage classification method reduces the amount of data processing and improves the classification speed and stability.

Keywords: Low-frequency oscillations; fuzzy matrices; three-stage random forest; machine learning; early warning and classification

1 Introduction

In recent years, under the premise of the vigorous development of smart grid and the proposed two-carbon target of China, the level of power grid has been continuously improved. The distance of power grid laid in China has reached a higher level. High voltage and long distance transmission make the stability problem of power system more and more prominent. The probability of low-frequency oscillation in power system is also greatly increased, which constantly threatens the safety of China's power grid. Once the problem of low-frequency oscillation in power system occurs, it will cause great losses to China's economy and people's lives. In order to ensure the safe operation of power grid, it is very important to study the early warning and classification of low frequency oscillation in power system.

The phenomenon of low-frequency oscillation in power system is caused by the interconnection of power grids. In the initial stage of power system interconnection, synchronous generators are closely connected, damping windings can generate enough damping, and low-frequency oscillation rarely occurs. With the expansion of power grid interconnection scale, and the wide adoption of high-magnification rapid excitation, the operation of power grid is getting closer and closer to the stability limit due to factors such as economy and environmental protection, and low-frequency oscillation in many power grids around the world. The problem of low-frequency oscillation has been paid more and more attention. Once the low-frequency oscillation in the form of amplitude occurs, it will seriously threaten the safe and stable operation of the power grid, and it is likely to induce a chain reaction accident, which will cause the stability of the system to be destroyed and make a large area of users lose power. Therefore, the study of low-frequency oscillation is of great significance.

Nowadays, the low-frequency oscillation of power system can be roughly divided into two types: local mode oscillation and inter-area mode oscillation. Generally speaking, the more units are involved and the wider the area is, the lower the oscillation frequency will be.

At present, in the domestic and foreign research on the early warning and classification of low-frequency oscillation in power system, Zhao Yan et al. proposed a two-stage Random Forest (RF) classifier [1] for the early warning and classification of low-frequency oscillation. However, the first-stage early warning part could not accurately give the early warning when considering the complex situation of multiple indicators. And the second stage classification still had the problem of complicated data volume. Yu Miao et al. used a method of low frequency oscillation amplitude warning indicator for power system based on a Vinnicombe distance of key feature wide-area downscaled data[2]. From the perspective of big data analysis, in practical experiments, the big data was transformed from a single time period matrix to a multi-time period high-dimensional matrix, and then transformed to a low-dimensional matrix. Finally, the study was carried out in combination with Vinnicombe distance. This method improved the accuracy of early warning to a certain extent, but it still relied on a single indicator for the early warning, and the possibility of misjudgment was large. Hu Hongmang et al. proposed a low-frequency oscillation mode recognition based on a wavelet packet and Support Vector Machine (SVM) [3], which solved the problem that the traditional low-frequency oscillation mode recognition method had a complex calculation process and poor real-time performance, that was difficult to meet the requirements of the classification and calculation of low-frequency oscillation modes in power system. However, this method still used a single indicator to judge the type of low-frequency oscillation. It still did not have an advantage in the comprehensive consideration of today's complex grid situation. Zhang Xiaohang et al. proposed a low-frequency oscillation type identification method based on an energy spatial-temporal distribution entropy [4], which could quickly and accurately judge the oscillation type and maintain the power grid security. But at present, this method was still a general classification method, and it could not judge the oscillation of specific region quickly. The power oscillation detection and classification based on a single node proposed by Steve Chan, Orlando et al. [5] showed that the system instability and power oscillation were initially interpreted as a completely connected K-part graph. But in practice, a special class of multi-part graph appeared. This special class multi-graph based on the numerical method identified the most likely types of oscillation as natural oscillation, transient oscillation and forced oscillation, but the method still had a large error in the classification of low frequency oscillation. Zhang Yuchen et al. investigated the reasons behind the misclassification of Extreme Learning Machine (ELM), and proposed a new Intelligent System (IS) [6] that could be released, which improved the traditional online Dynamic Safety Assessment (DSA) was not enough to meet the needs of online DSA for the gradual simulation of the power system dominated by the renewable energy when encountering a large number of emergency sets or many operating parameters. Zheng Wei et al. proposed a boundary-based research on low-frequency oscillation early warning and control strategy of wind power system [7], and also proposed a new indicator of low-frequency oscillation early warning. In the calculation of low-frequency oscillation early warning, this new type of indicator not only simplified the calculation method, but also enhanced the possibility of complex calculation, and provided a new idea for low-frequency oscillation early warning. The framework of "double-high" power system broadband oscillation wide-area monitoring and early warning system proposed by Ma Ningning et al. [8] could realize the monitoring and early warning of broadband oscillation at the same time as the early warning of low-frequency oscillation. The system substation could realize the acquisition and calculation of fundamental wave, harmonic and interharmonic phasors, and finally realized the protection control of broadband oscillation, so as to eliminate broadband oscillation and reduce the risk of system operation. Alishba Sadiq et al. proposed a classified Alzheimer's disease using low-frequency fluctuations of rs-fMRI signals [9], and the proposed classification could apply to the detection of low-frequency oscillations in power systems that also detected low-frequency fluctuation signals. Improvements of classification metrics and systems could also result in higher accuracy. The distinction and conversion oscillation mode and rotor angle mode proposed by He Tingyi et al. [10] solved the problem that the coherent oscillation of generator speed was the key feature of distinguishing frequency oscillation and power angle oscillation in existing research. Du weijie et al. proposed Research on low-frequency oscillation safety early warning and situational awareness prediction mechanism of new power system[11], which can reduce the influence of excessive subjectivity due to expert evaluation matrix, and can effectively solve the problem of low-frequency oscillation early warning of new power system with multiple indicators.

In the early warning of low-frequency oscillation in power system, the traditional problems are still the single early warning indicator and the serious problem of low early warning accuracy. The accuracy of early warning can be greatly improved by introducing comprehensive indicators for the comprehensive analysis. In the classification of low-frequency oscillation, the main problem of traditional methods is that the classification speed is slow due to the large amount of data. The classification indicator is single, and some cases are ignored, so the classification effect is not ideal. To solve the problem of data processing, it is mainly to propose more classification methods in other fields to classify the low-frequency oscillation phenomenon. In terms of classification indicators, the main solution is put forward to using more indicators for more detailed classification, more types, and the effectiveness of classification.

The contribution of this paper is as follows.

(1) In the first stage of Random Forest (RF), the "fuzzy judgment matrix" is introduced to

reduce the problem of one-sided judgment caused by the single indicator and greatly improve the accuracy of early warning of low-frequency oscillation.

(2) After judging the low-frequency oscillation of the power system has entered the second stage, the second judgment is made. Reducing the amount of data processing will improve the classification speed and stability to some extent. It ensures the speed and stability of RF in practical application.

The rest of this paper is organized as follows. Section 2 gives the basic theory of fuzzy matrix and SWT transformation. Section 3 begins with the stochastic forest theory and describes the improved three-stage stochastic forest. Section 4 selects the data of IEEE 39 nodes to simulate the first stage and selects the data to simulate the second and third stages of the random forest, and compare the results with the traditional second-order random forest to get better results. Section 5 summarizes the whole paper.

2 Fundamental theory

2.1 Fuzzy evaluation matrix

It is difficult to fuse several indicators, because the data measured in Phasor Measurement Unit (PMU) have different physical properties, so it is difficult for traditional methods to fuse multiple indicators to judge and warn the occurrence of low-frequency oscillation. According to the relationship between different indicators, we use the Kernel matrix [12] in Fuzzy Analytical Hierarchy Process (FAHP) to calculate the correction degree between elements in fuzzy judgment matrix. This paper finds out a research mechanism of low-frequency oscillation early warning of comprehensive indicators, which is used in RF 1 stage to improve the accuracy of low frequency oscillation early warning.

(1) Consistency check

According to the expert evaluation matrix, a kernel matrix is introduced and the consistency test is carried out.

Consistency indicator h:

$$h = \frac{2}{n(n-1)(n-2)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \sum_{\substack{k=1 \\ k\neq i,j}}^{n} \left| u_{ker}^{m} - (u_{ker}^{m} + u_{ker}^{m} - \tfrac{1}{2}) \right| \tag{1}$$

(2) Fuzzy evaluation matrix

When the fuzzy comprehensive evaluation of the low-frequency oscillation safety early warning indicator of the new power system is carried out, there is a fuzzy demarcation domain in the operating state displayed by the early warning indicator. The amplitude membership function, frequency membership function, peak-peak membership degree function and power oscillation first pendulum peak membership function of each indicator are obtained [13].

The fuzzy evaluation matrix R is obtained by substituting each early warning indicator by online identification calculation into the membership function.

$$R = \begin{bmatrix} \varphi_{11} & \varphi_{12} & \cdots & \varphi_{1n} \\ \varphi_{21} & \varphi_{22} & \cdots & \varphi_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{m1} & \varphi_{m2} & \cdots & \varphi_{mm} \end{bmatrix} \tag{2}$$

ISSN: 0028-0712

$$K(u_{ij}^m) = \begin{cases} a_{ij}^m, a_{ij}^m = b_{ij}^m = c_{ij}^m = d_{ij}^m \\ \frac{(c_{ij}^m)^2+(d_{ij}^m)^2-(a_{ij}^m)^2-(b_{ij}^m)^2+d_{ij}^m c_{ij}^m-a_{ij}^m b_{ij}^m}{3(c_{ij}^m+d_{ij}^m-a_{ij}^m-b_{ij}^m)} \end{cases} \quad (3)$$

(3) Safety degree of low-frequency oscillation early warning in power system

According to the established weight vector W and fuzzy evaluation matrix R, the membership degree C of each early warning indicator under three levels is calculated.

$$C = W \otimes R = (b_1, b_2, b_3) \rightarrow S = e^{5b_3+3b_2+b_1} \quad (4)$$

In formula (4), ⊗ is a fuzzy operator symbol, which is commonly used operator symbols including a weighted average operator, and a Zadeh operator, etc. The fuzzy comprehensive evaluation model will vary according to the operator. In order to make the algorithm more accurate, the Zadeh operator is selected to obtain the comprehensive early warning indicator S of the new power system.

The Zadeh operator（∨，∧）is a large and small operation on membership[14].Taking the first column of elements W and R as an example, we compare the size of the first column of teammate elements in W and R, find the maximum value, and then take the smaller of the obtained maximum value.

We can get the comprehensive warning indicator S safety level:

Table 1 Comprehensive warning indicator *S* safety level

Safety Level	Safety	Danger	High Risk
S	＜0.43	0.43~0.72	＞0.72

Finally, we get the comprehensive judgment indicator S, and use S to judge whether the system has low frequency oscillation or not.

2.2 SWT theory

Synchrosqueezing Wavelet Transform (SWT) is based on Continue Wavelet Transform (CWT), which uses the characteristic that the phase of signal in frequency domain after the wavelet transform is not affected by scale transformation to find the corresponding frequency at each scale[15]. Then add the scales at the same frequency, that is, redistribute the wavelet coefficients obtained by wavelet transform and compress them, so that the values near the same frequency are compressed into this frequency, which improves the ambiguity in the scale direction and the time-frequency resolution. In order to make the RF classifier process more simple and smart, it is necessary to identify the parameters of low-frequency oscillation signals by the SWT method based on the data obtained from the actual PMU.

(1) The obtained low frequency oscillation signal $x(t)$ is transformed by CWT.

$$W_x(a,b) = \int_{-\infty}^{+\infty} x(t) a^{-1/2} \overline{\psi}\left(\frac{t-b}{a}\right) dt \quad (5)$$

In formula (5), a is a scale factor and b is a translation factor, which is also known as a time translation factor. $\overline{\psi}(\frac{t-b}{a})$ is a conjugate wavelet base, which is generated by the translation and expansion of the basic wavelet function.

(2) Divide the sampling interval. Assuming the length of signal $x(t)$ is $n = 2^{L+1}$ and sampling interval is Δt. Therefore, the frequency of low-frequency signal can be divided into several different

frequency intervals according to Nyquist sampling theorem:

$$W = [\frac{\omega_{l-1}+\omega_l}{2}, \frac{\omega_l+\omega_{l+1}}{2}] \tag{6}$$

(3) Transform the magnitude of the discrete wavelet.

$$T_x(\omega_l, b) = (\Delta\omega)^{-1} \sum_{a_k:|\omega_x(a,b)-\omega_l| \leq \frac{\Delta\omega}{2}} W_x(a,b) a_k^{-\frac{3}{2}} (\Delta a)_k \tag{7}$$

(4) Extract modal components of the low frequency oscillation. The first component can be reconstructed by formula (8) The $x_k(t)$ component of $x(t)$ can be reconstructed using formula (3).

$$x_k(t_m) \approx \frac{2}{R_\psi} Re[\sum_{l \in L_k(t_m)} \tilde{T}_{\tilde{x}}(\omega_l, t_m), \tag{8}$$

In formula (8), $L_k(t_m)$ is the subscript set near the curve ω_l (the k-th IMT is component of $x_k(t)$); $t_m$is the discrete value of t.

(5) The modal parameters of low-frequency oscillation are calculated by the Hilbert (HT) transform. In the general control theory, the oscillation characteristic of damping oscillation link is as follows:

$$x(t) = Ae^{\lambda t} cos(\omega t + \rho_0) = Ae^{-\xi\omega_0 t} cos(\omega_0\sqrt{1-\xi^2 t} + \rho_0) \tag{9}$$

In formula (9), λ is an attenuation factor; ξ is a damping ratio; ω is a undamped frequency. Therefore, it can be concluded that:

$$\begin{cases} \xi\omega_0 = \lambda \\ \omega_0\sqrt{1-\xi^2} = \omega \end{cases} \tag{10}$$

It can be obtained from formula (9) that the calculation formula of damping ratio is

$$\xi = \frac{\pm\lambda}{\sqrt{\omega^2+\lambda^2}} \tag{11}$$

3 Classification of Three-stage Random Forest in Low Frequency Oscillation

3.1 Classification and Random forest

Classification and random forest is a classifier that uses multiple trees to train and predict samples, which belongs to the category of machine learning [16]. On the premise of having some historical data as training set, it can randomly select training data for machine learning, and then automatically classify the data we need to process.

The algorithm steps of RF are as follows:

Step1: First, determine the number of training cases (samples) and test samples, and determine the number of features.

Step2: From the training cases (samples), samples are taken for many times in the way of putting back samples to form a training set. Cases (samples) have not been drawn to make predictions and evaluate their errors.

Step3: Machine learning the training set.

Step4: After the machine learning, RF has the ability to classify, and the remaining test sets are

ISSN: 0028-0712

tested. After classification, the accuracy of the RF is obtained by comparing with the correct results.

The RF structure diagram we used is shown in Figure 1.

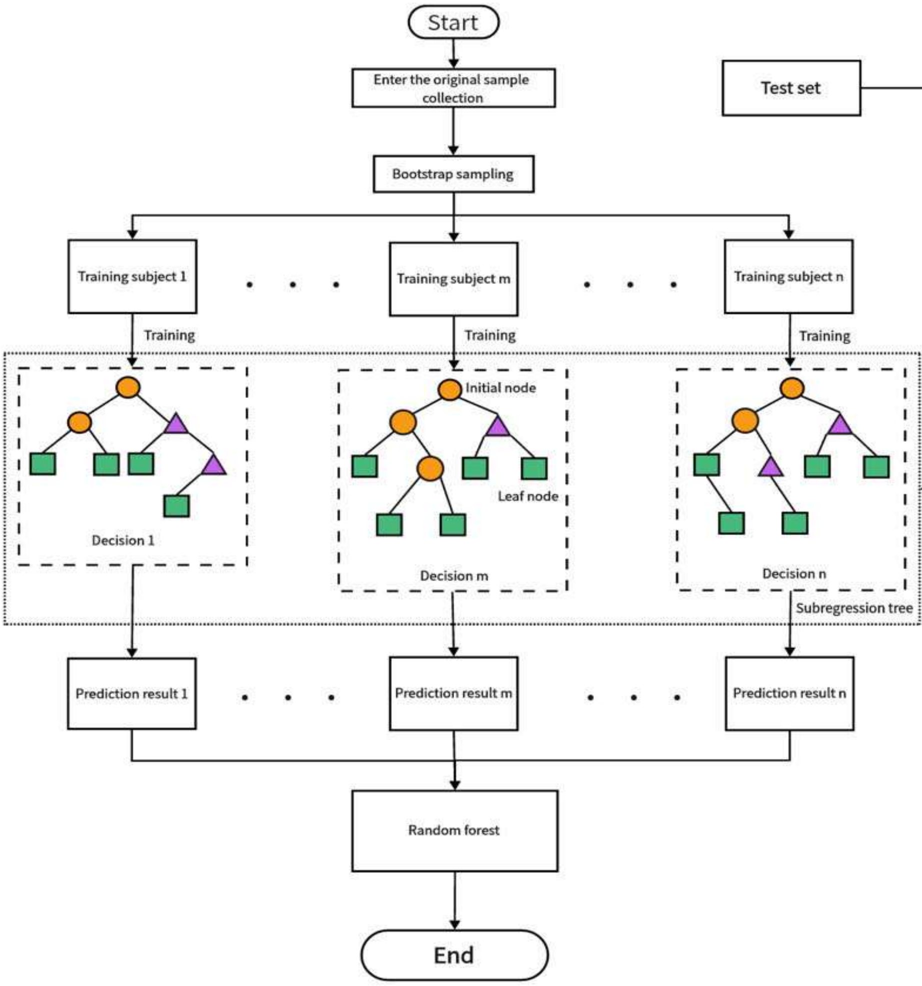

Figure1 Classification and Random forest process

3.2 Three-stage Random Forest

3.2.1 Random forest 1

The main purpose of the first random forest classifier is to judge whether the data has low-frequency oscillation and give timely warning when low-frequency oscillation occurs.

Firstly, we judge S obtained by the fuzzy matrix fusion:

Category 1:$S \geq 0.43$, which means that the data of this section has a high probability of low-frequency oscillation and will enter the next stage for the judgment.

Category 2: $S < 0.43$, which means that there is no low-frequency oscillation in this section of data.

3.2.2 Random forest 2

The second random forest classifier mainly determines the low-frequency oscillation of forced oscillation type and makes a judgement for λ.

Enter this stage when it is judged as category 1 at the first stage. At this time, we will use the damping ratio and attenuation factor obtained after the SWT transformation and HT transformation for the second judgment:

Category 1: $\lambda < 0$, the judgment result is the forced oscillation.

Category 2: $\lambda \geq 0$, if the judgment result is weak damping oscillation or negative damping oscillation, it will go to the third stage for the final judgment.

3.2.3 Random forest 3

The third random forest is the final decision. At this stage, ξ is judged and the types of weak damping oscillation or negative damping oscillation are divided.

Enter this stage when the second stage is category 2, and we will make the final classification:

Category 1: $\xi < 0$, and it is judged that negative damping oscillation occurs.

Category 2: $\xi > 0$, and it is judged that weakly damping oscillation occurs.

3.3 Classification process

Three-stage random forest process steps are as follows.

Step 1: Fuzzy matrix construction

The measured PMU data is evaluated and analyzed by the fuzzy matrix, and a comprehensive indicator score S is obtained.

Step 2: The first stage judgment

We use comprehensive indicator to judge whether low-frequency oscillation occurs. From Table 1, S is compared with the value of 0.43, and if S is less than 0.43, the data in this section is normal. If S is greater than or equal to 0.43, the low-frequency oscillation will occur in this section of data, and early warning will be given and the classification of low-frequency oscillation in the next stage will be started at the same time.

Step 3: SWT analysis

SWT processes the data with low frequency oscillation for data analysis of RF 2 and RF 3. After the data processing, the processed data is tested once, and the problem data with $\xi < 0$ and $\lambda < 0$ is deleted to ensure the normal classification of RF.

Step 4: The second stage judgment.

The historical PMU data is used as the training set, and the sample training is also carried out after SWT transformation. The measured data is used as the test set, and RF 2 pairs are used to judge. If $\lambda < 0$, it is judged as forced oscillation. If $\lambda > 0$, it is judged as weak damping oscillation or negative damping oscillation, and the third stage is judged.

Step 5: The third stage judgment.

RF 2 and RF 3 also transform the historical PMU data, conduct the sample training to test the measured data and judge ξ. If $\xi < 0$, then negative damping power oscillation occurs, and if $\xi > 0$, weak damping oscillation occurs.

According to the above steps, we can get the Figure 2 as follows.

ISSN: 0028-0712

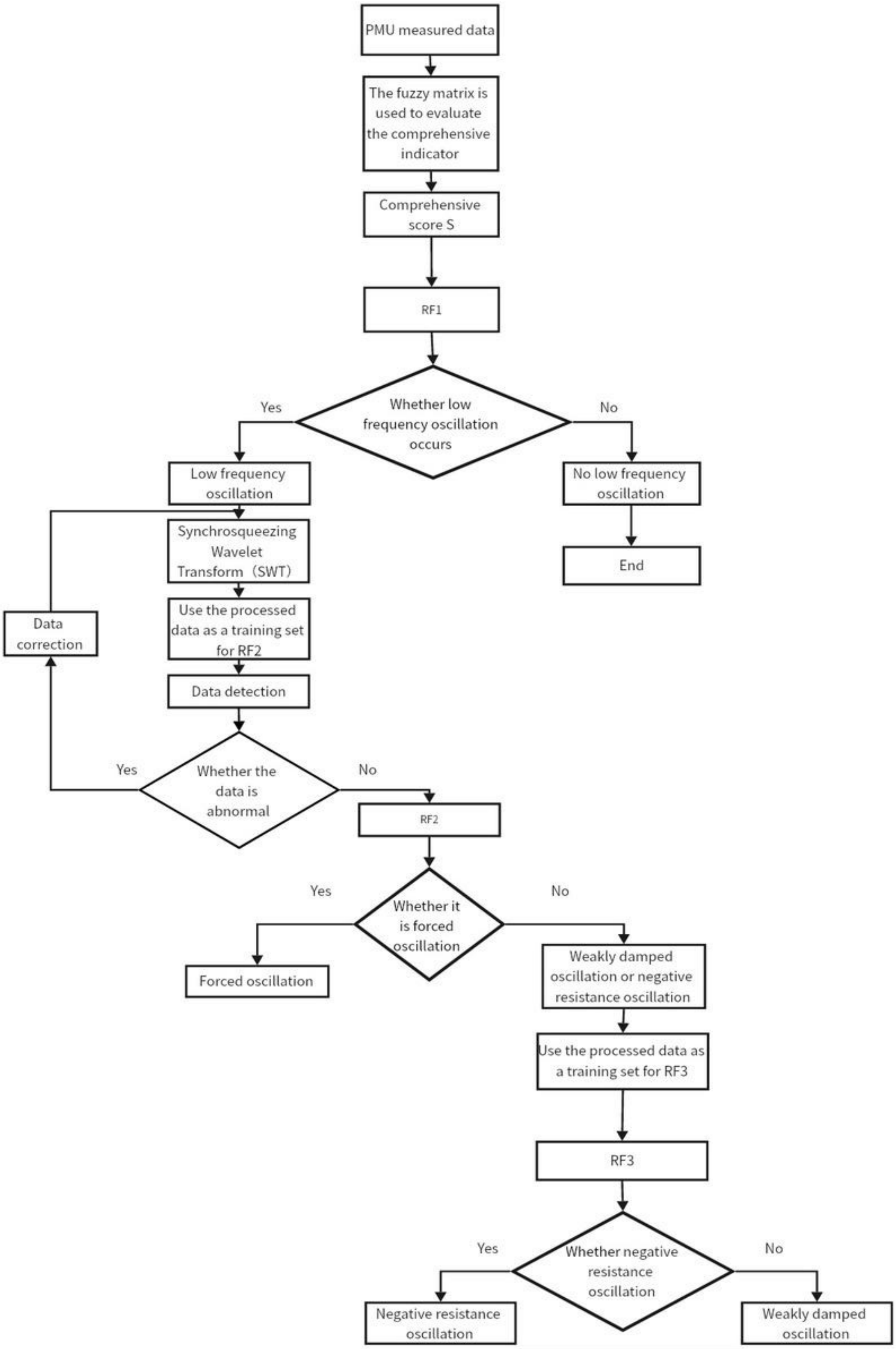

Figure 2 Three-stage random forest process

ISSN: 0028-0712

4. Simulation and Verification

4.1 New England 10-machine and 39-node System

In order to verify the feasibility and effectiveness of the method proposed in this paper, we use the New England 10-machine and 39-node system, which is a typical system as a simulation validation based on the principle of real PMU data distribution used in [17], which is as shown in Figure 3.

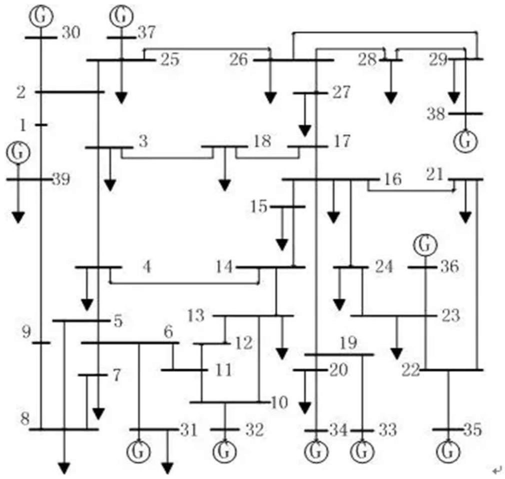

Figure 3 New England 10-machine and 39-node system

4.2 Test data and results of RF 1

First, we can try to compare the results of low-frequency oscillation which uses a single frequency with the results of low-frequency oscillation phenomena judged by comprehensive indicators. We read and analyze a PMU data from 0 seconds to 180 seconds. And then we can get the following pictures of frequency, current and voltage. Figure 4 shows the frequency characteristics of data monitored by PMU within 0 seconds~180 seconds.

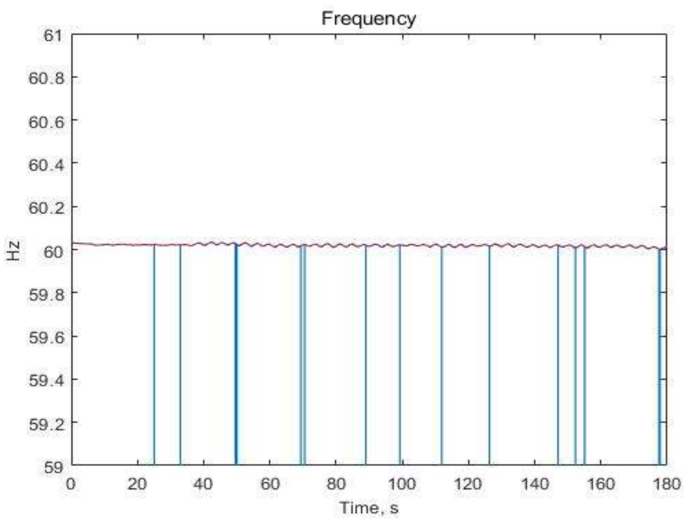

Figure 4 Frequency data

As can be seen from Figure 4, the frequency of this data segment is stable within 0 seconds~180

ISSN: 0028-0712

seconds, and no obvious abnormality occurs. According to the Frequency of Figure 4, it is judged that it does not have low-frequency oscillation.

Figure 5 and Figure 6 show the characteristics of current and voltage detected by the PMU data within 0 seconds~180 seconds, respectively.

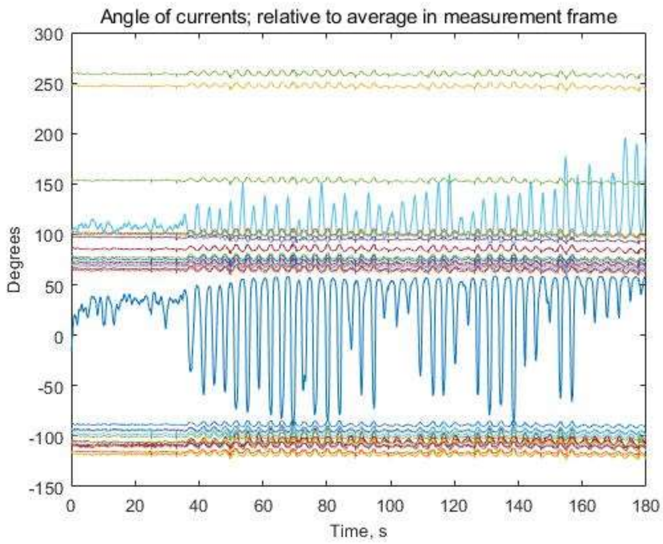

Figure 5 Current data

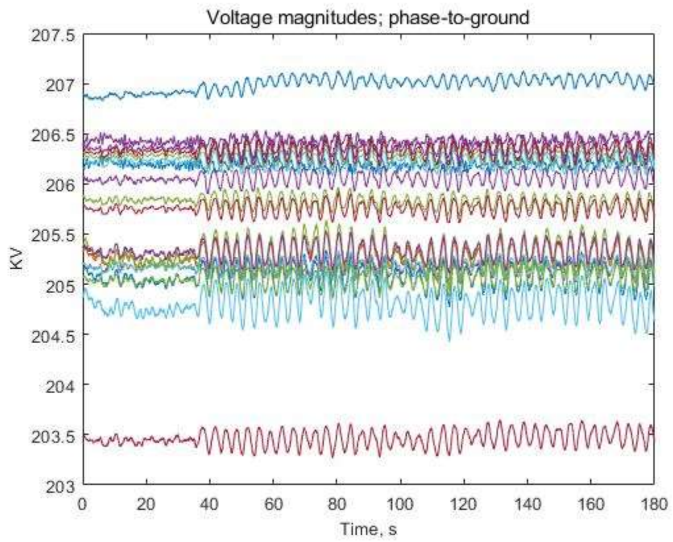

Figure 6 Voltage data

According to the current and voltage data, there is obvious low-frequency oscillation in this period of data around 40 seconds ~180 seconds, which proves that it is very limited to judge the occurrence of low-frequency oscillation only by frequency, and it is necessary to introduce comprehensive indicators to consider the occurrence of low-frequency oscillation.

ISSN: 0028-0712

4.3 Test results by introducing fuzzy matrix:

(1) Determine the weight of each indicator:

$$U=\begin{bmatrix}(0.5,0.5,0.5,0.5) & (0.3,0.5,.5,0.7) & (0.7,0.9,1,1) & (0.8,1,1,1)\\ (0.3,0.5,0.5,0.7) & (0.5,0.5,0.5,0.5) & (0.6,0.8,0.8,1) & (0.7,0.9,1,1)\\ (0,0,0.1,0.3) & (0,0.2,0.2,0.4) & (0.5,0.5,0.5,0.5) & (0.6,0.8,0.8,1)\\ (0,0,0,0.2) & (0,0,0.1,0.3) & (0,0.2,0.2,0.4) & (0.5,0.5,0.5,0.5)\end{bmatrix}$$

$$A=\begin{bmatrix}1 & 3 & 5 & 7\\ 1/3 & 1 & 3 & 5\\ 1/5 & 1/3 & 1 & 3\\ 1/7 & 1/5 & 1/3 & 1\end{bmatrix}$$

Weights are obtained by the expert evaluation matrix w_U:

$$w_U=[0.3533 \quad 0.3347 \quad 0.2042 \quad 0.1077]^T$$

Weights are obtained by the expert evaluation matrix w_A:

$$w_A=[0.5650 \quad 0.2622 \quad 0.1175 \quad 0.0553]^T$$

Select a set of data between 60 seconds and 80 seconds, then we can get:

$$R=\begin{bmatrix}1.0 & 0.5 & 1.0 & 0.5\\ 1.0 & 1.0 & 1.0 & 0.5\\ 0.5 & 0.6 & 1.0 & 0.8\\ 0.9 & 0.1 & 0.9 & 0.8\\ 0.5 & 0.7 & 1.0 & 0.7\end{bmatrix}$$

W is obtained by fuzzy matrix calculation:

$$W=[0.9357 \quad 0.6207 \quad 0.9945 \quad 0.5518]$$
$$C=W\circ R=(b_1,b_2,b_3)$$
$$S=e^{5b_3+3b_2+b_1}$$

Then, the final calculation S is 0.57.

Obviously, this section of data S>0.43, which proves that low-frequency oscillation occurs on this section of data selected from 60 seconds to 80 seconds, so as to make early warning and enter the next stage for judgment.

(2) Select a set of data from 0 seconds to 20 seconds, then we can get R:

$$R=\begin{bmatrix}0.3 & 0.2 & 0.4 & 0.5\\ 0.2 & 0.1 & 0.3 & 0.4\\ 0.1 & 0.2 & 0.3 & 0.1\\ 0.5 & 0.3 & 0.4 & 0.7\\ 0.3 & 0.5 & 0.5 & 0.6\end{bmatrix}$$

w is obtained by fuzzy matrix calculation:

$$w=[0.2613 \quad 0.1793 \quad 0.3620 \quad 0.4378]$$

ISSN: 0028-0712

$$C = W \circ R = (b_1, b_2, b_3)$$
$$S = e^{5b_3 + 3b_2 + b_1}$$

Then, the final calculation S is 0.22.

It can be found that S<0.43 from the data is selected from 0 seconds to 20 seconds, indicating that there is no low frequency oscillation from this data.

The results of RF 1 are consistent with Figure 4, which proves the feasibility of RF 1 stage.

4.4 Test data and results of RF 2

The attenuation factor and damping ratio of a piece of data are obtained by the SWT post-transformation, and we can get the attenuation factor and the damping ratio. Then we use the machine learning process to classify the data and get the Table 2, "1" means forced oscillation and "2" means weakly damped oscillation.

Table 2 RF 2 test set

Attenuation Factor	Damping Ratio	Category
-0.265	0.232	1
-0.123	0.322	1
-0.678	0.312	1
-0.453	0.123	1
-0.123	0.123	1
-0.212	0.123	1
-0.877	0.123	1
-0.877	0.123	1
-0.472	0.869	1
-0.321	0.765	1
-0.542	0.309	1
0.233	-0.756	2
0.098	-0.326	2
0.438	-0.385	2
0.123	-0.073	2
0.234	-0.113	2
0.234	-0.463	2
0.334	-0.471	2
0.368	-0.673	2
0.638	-0.722	2
0.678	-0.323	2
0.256	-0.534	2
0.234	-0.542	2

0.678	-0.324	2
0.877	-0.653	2
0.877	-0.653	2
0.871	-0.453	2
0.856	-0.543	2
0.368	0.342	2
0.375	0.433	2
0.633	0.223	2
0.234	0.323	2
0.567	0.324	2

From Table 2, we can know that 80% of RF2 data is trained, and the remaining 20% data is judged and verified. The accuracy of the final test results reaches 100%, which shows that the learning discrimination of random forest at RF2 stage is correct and proves the feasibility of RF2 stage.

4.5 Test data and results of RF 3

SWT transformation of RF 3 is carried out in the same way to obtain the data and then machine learning of that data is done. Table 3 shows us the RF 3 test set.

Table 3 RF 3 test set

Attenuation Factor	Damping Ratio	Category
0.233	-0.756	1
0.098	-0.326	1
0.438	-0.385	1
0.123	-0.073	1
0.234	-0.113	1
0.234	-0.463	1
0.334	-0.471	1
0.368	-0.673	1
0.638	-0.722	1
0.678	-0.323	1
0.256	-0.534	1
0.234	-0.542	1
0.678	-0.324	1
0.877	-0.653	1
0.877	-0.653	1
0.871	-0.453	1
0.856	-0.543	1
0.368	0.342	2
0.375	0.433	2
0.633	0.223	2
0.234	0.323	2
0.567	0.324	2
0.789	0.223	2
0.243	0.223	2
0.253	0.123	2

0.352	0.112	2
0.326	0.232	2
0.256	0.212	2
0.423	0.213	2
0.243	0.431	2
0.987	0.312	2

The same proportion of training and testing is carried out for RF 3, and the accuracy of the final test result is 100%, which proves the feasibility of RF 3 stage.

We have tested a total of 1000 sets of data and finally obtained classification results, in which 765 sets are forced oscillations, 134 groups are weakly damped oscillations, and 101 groups are negative resistance oscillations. We also analyze the possible causes of various types of oscillations, as shown in Table 4.

Table 4 Classification results

Type of Oscillation	Number of Occurrences	Occurrence Reason
Forced oscillation	765	There are devices or lines in the system with mismatched frequency characteristics
Weakly damped oscillation	134	In an unstable state, there is a source of interference
Negative resistance oscillation	101	A single instantaneous interference causes power oscillations within the system that do not subside

4.6 Data comparison

We use the data from Table 4 to make scatter plots. Then we can see in Figure 7 that red dots are in a separate area and can be easily separated out.

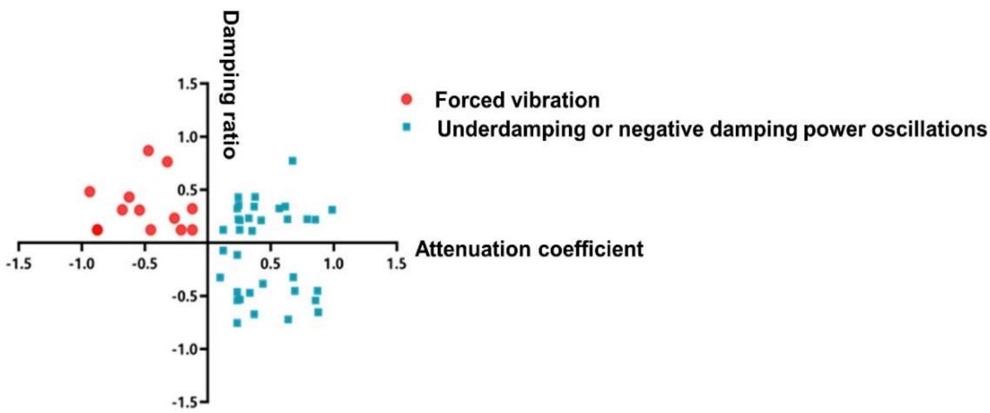

Figure 7 RF 2 test results

In this way, we can easily remove some of the data and then classify the remaining data. Then we can get the test results from Figure 8.

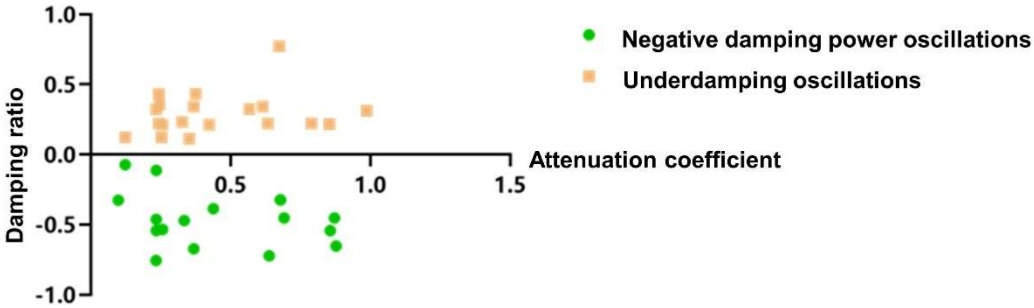

Figure 8 RF 3 test results

We assume that there are 2000 groups of data with low frequency oscillation to be analyzed, in which forced oscillation accounts for 60%, negative damping power oscillation and weak damping oscillation accounts for 20% respectively. Only the two-stage random forest of RF 2 needs to process 4000 times of classification and discrimination. A three-stage random forest with RF 3 needs 3600 times of classification and discrimination. Suppose there are N groups of low frequency oscillation, and the forced oscillation accounts for $M\%$.

The number of calculations required for the two-stage is:

$$N \times 2$$

The number of calculations required for the three-stage is:

$$N + (N - N \times M\%) \times 2 = 3 \times N - N \times 2 \times M\% = (3 - 2 \times M\%) \times \ N$$

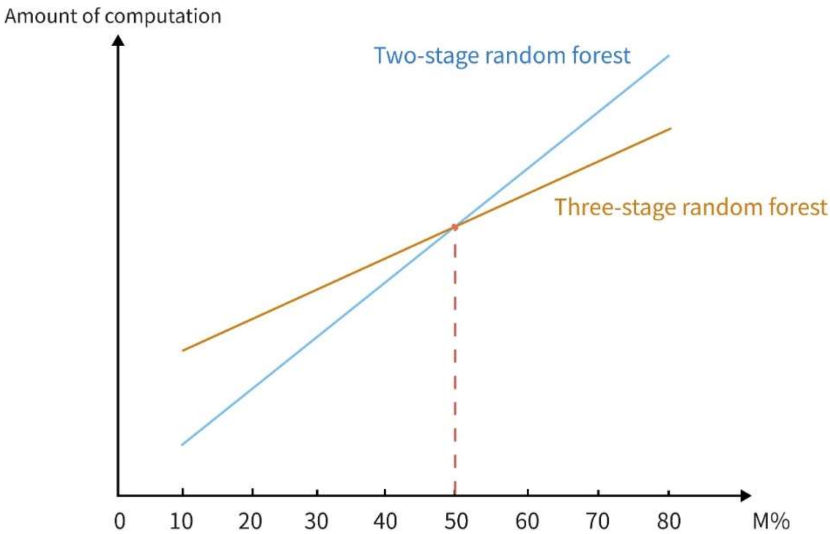

Figure 9 Comparison of the two-stage RF calculation and the three-stage RF calculation

It can be seen from Figure 9 that the two-stage RF and the three-stage random forest reach the same number of calculations when the M value is 50%, and the number of calculations of the three-stage random forest will continue to widen the gap with the two-stage RF, which can above 50%. In the red experiment, more than 30 sets of PMU data can be obtained per second, and the amount of data to be processed is far greater than 2000 data sets. A smaller amount of processing will greatly improve the classification speed, which will ensure the stability of the processor and improve the working ability of the processor.

With the large number of power system stabilizers putting into use, the low-frequency oscillation of negative damping mechanism has been greatly reduced, while the low-frequency oscillation of forced oscillation mechanism has occurred more and more. The higher proportion of forced oscillation represents the further reduction of data processing, and the advantages of the three -stage RF will become more obvious.

5. Conclusion

The classification of low frequency oscillation is of great significance to the study of low frequency oscillation early warning. In this paper, aiming at some shortcomings of the two-stage RF, a three-stage RF is improved and optimized respectively.

(1)The fuzzy evaluation matrix is introduced at the RF 1 stage, and the comprehensive evaluation indicator S is calculated to replace the original single indicator frequency. It solves the problem that a single indicator frequency may not be able to predict the low-frequency oscillation that may occur in the complex power grid and improves the accuracy of prediction.

(2)At the RF 2 stage, the original two-stage RF is improved into a three-stage RF, which optimizes the data processing capacity, improves the speed of classification, ensures the stability of the system operation under the big data, and caters the high demand of forced oscillation in reality.

Compared with the two-stage RF, the three-stage RF solves the problem of large data processing in the power system. The method proposed in this paper can effectively solve some existing problems in the early warning and classification of low-frequency oscillations in power system and has a good reference value for the early warning and classification of low-frequency oscillations in large-scale power systems. In the future, we will continue to optimize the technology for the early warning and classification direction of low-frequency oscillation and improve the overall accuracy and efficiency by timely adjusting the prediction indicator and classification indicator, so as to protect the safe operation of the power system.

6.References

[1] Zhao Y, Huo H, Xu H. Application of two-stage random forest classification method in low-frequency oscillation monitoring [J]. Journal of Northeast Electric Power University, 2020, 40(2):60-67.

[2] Yu M, Du W J. Research on early warning accuracy of low frequency oscillation in power system based on Vinnicombe distance [J]. Control Engineering, 2022,29(02):339-347.

[3] Hu H, Xiao H, He X, et al. Low-frequency oscillation pattern recognition based on wavelet packet and SVM[C]// The 3rd Conference on Energy Internet and Energy System Integration (EI2). 8-10 November, 2019,Changsha, China.pp:1442-1446.

[4] Zhang X H, Zhang W Z, Xi J H, et al. Discrimination method of low-frequency oscillation type based on energy spatio-temporal distribution entropy[J]. Power System Automation,2017,41(08):

84-90.
[5] Steve C, Parnmook N. Single node-based power oscillation detection and classification[C]// The 8th International Conference on Condition Monitoring and Diagnosis (CMD), 25-28 October,2020, Phuket, Thailand.pp:182-185.
[6] Zhang Y, Yan X, Zhao Y D, et al. Intelligent early warning of power system dynamic insecurity risk: Toward optimal accuracy-earliness tradeoff[J]. IEEE Transactions on Industrial Informatics, 2017,13(5):2544-2554.
[7] Zheng W, Ge R D, Guo P, et al. Study on analysis and control strategy of power system with wind power based on the boundary[J]. China Electric Power (Technical Edtion), 2016, (03):1-6.
[8] Ma N N, Xie X R, Tang J, et al. Wide-area monitoring and early warning system of broadband oscillation of "double-high" power system[J]. Journal of Tsinghua University: Natural Science Edition, 2021,61(5):457-464.
[9] Shi L, Gu W, Fan C, et al. Identification of low frequency oscillation mechanism based on deep learning[C]//The 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), 20-23 Sept, 2020, Nanjing, China. pp:1-6.
[10] He T Y, Li S N, Cen L .Distinction and conversion between frequency oscillation mode and rotor angle oscillation mode[C]//The 5th IEEE Conference on Energy Internet and Energy System Integration (EI2), 22-25 Oct, 2021, Taiyuan, China.pp:371-375.
[11] Du W J. Research on low-frequency oscillation safety early warning and situation awareness prediction mechanism of new power system[D].Beijing University of Civil Engineering and Architecture,2023.
[12] Wang C, Li B, Jiang T, et al. A contingency screening and ranking method based on improved fuzzy analytic hierarchy process for AC/DC receiving-end power grid AC faults[J]. Power System Technology,2021,45(10): 4047-4056.
[13] Yu M, Du W J, Li J L, et al. Study on a comprehensive indicator and online classification of early warning of low frequency oscillation in power system[J].Engineering Applications of Artificial Intelligence, 2023, 124:106592.
[14] Ran C,Kunhua J,Yun W, et al. Management Evaluation of Distribution Network Based on Fuzzy Evaluation Matrix[J]. Journal of Physics: Conference Series,2019,1345(2):1-6.
[15] Liu Y, Li K, Chen P. Fault diagnosis of rolling bearing based on synchronous compression wavelet transform [J]. China Mechanical Engineering, 2018, 29(5):585-590.
[16] Liu L, Zeng J. Design and application of a combinatorial classification algorithm based on random forest[J]. Electronic Design Engineering,2020,28(16):54-57.
[17] Maslennikov S, Wang B, Zhang Q, et al. A test cases library for methods locating the sources of sustained oscillations[C]//IEEE PES General Meeting, 17-21 July,2016, Boston, MA.pp:1-5.