Research and application of a real-time operation risk assessment method for power grid based on random forest algorithm

CHEN Junquan, WU Shihong, YE Hangchao, YANG Fu, SONG Xian
Guizhou Power Grid power dispatch and control centre,
Jiefang Road #86, Nanming district, Guiyang, Guizhou province, China

E-mail: 879482999@qq.com

Abstract. Real-time operation risk assessment of power grid is not only an important content of power grid dispatching and operation control, but also an important reference factor in power grid accident disposal. It is difficult for the traditional evaluation methods to meet power system online operation control requirements, because of their complex process and long calculation time. To this end, a real-time power grid operation risk assessment method based on random forest intelligent algorithm is proposed in this paper. The basic concepts and implementation points of random forest algorithm is introduced. The implementation framework of real-time operation risk assessment based on random forest is proposed taking the need of power grid operation risk assessment into consideration. Compared with the traditional evaluation method, this method can calculate the power grid operation risk level directly according to some key operation indexes without complex power grid analysis, such as N-1 security check. Finally, the two-years practical application in a certain provincial power grid in China shows that the method has high evaluation accuracy and computational efficiency, which can match the traditional evaluation method and improve the calculation efficiency and accuracy of risk assessment.

1. Introduction

The core idea of traditional power grid operation risk assessment method is to modelling the power grid operation, which has three main steps, including: (1) calculating the failure probability of power grid transmission and transformation equipment, by considering different meteorological conditions; (2) assessing the accident impact of different equipment failures; (3) counting the accident consequences of all faults and calculating power system operation risk.

A stochastic fuzzy evaluation model of equipment failure based on the credible theory is constructed by considering the influence of weather and other factors on equipment reliability, which is used to propose a risk assessment method for power grid operation in ref. [1]. The ref. [2-5] focuses on the influence of large-scale wind power on power grid operation. Then a risk assessment method of power grid operation under large-scale wind power connection is proposed. The ref. [6] proposed a risk assessment method for small interference stability, by considering the stability characteristics of small interference in large power grid. The ref. [7] proposes a hierarchical intelligent evaluation system for evaluating voltage stability characteristics of large power grid. The ref. [8] studies the action strategy of protection device and puts forward a dynamic assessment method of power grid operation risk considering equipment protection action, which can be used to simulate the development process of power system device failures. A power grid situational awareness and economic operation control
system based on operation risk assessment is proposed, which provides reference for power grid operation control based on multi-period operation risk analysis in ref. [9].

The above research provides useful references for the power system real-time operation risk assessment. With the expansion of power grid scale, the complexity of power system operation is increasing rapidly, which causes an urgent need for an efficient power grid operation evaluation method. However, it is difficult for the traditional evaluation methods to meet power system online operation control requirements, because of their complex process and long calculation time.

To this end, this paper will propose a real-time power system operation risk assessment method based on random forest. Firstly, the basic concept and implementation of the method will be introduced. Then, the decision-making framework of this method would be proposed. Finally, the application of the method in a provincial power grid is introduced, including system development and application benefit.

2. Basic concept and implementation framework of stochastic forest algorithm

2.1. Basic concept

The so-called random forest is essentially an integrated machine learning agent composed only of decision trees. Its basic idea is to build multiple decision trees with strong independence to obtain the final decision results through comprehensive analysis, to improve the accuracy of the overall decision analysis. Each decision tree that constitutes a random forest is called a base learner. Compared with other integrated machine learning algorithms, random forest has higher generalization error and training efficiency, which means that the random forest algorithm can obtain more independent base learners in a shorter time, so that it can solve practical problems with higher complexity more effectively and avoid falling into local optimal solutions. It has been proved that the upper bound of generalization error of random forest can be expressed as follows:

\[ p_{\text{max}} = \frac{\sigma(1-\rho)}{s} \]  

in which \( p_{\text{max}} \) is the upper limit of generalization error of the random forest, \( \rho \) is the average correlation coefficient between decision trees of the random forest and \( s \) is the classification intensity between decision trees. According to glass database and auto-mpg database, the monitoring error rate of random forest agent is about 4% higher than that of other Bagging integrated learning intelligent algorithms.

2.2. Implementation framework

As shown in figure 1, when training to generate random forest agents using training data sets, it mainly includes four steps: data pre-processing and modeling, setting the subset attribute number, decision tree training and summary of decision results.

![Training points of random forest agents](image)

**Figure 1.** Training points of random forest agents

(1) data pre-processing and modeling

The main task of this step is to eliminate the bad data in the training data set according to the actual problems to be solved, complete the possible missing data items, and convert the attributes that do not meet the requirements of data types into discrete attribute values.
(2) setting the subset attribute number
The main task of this step is to set the attribute value of each decision tree for the subsequent decision tree training. Compared with the traditional decision tree agents, the biggest difference of random forest lies in that every step of the traditional decision tree training generation is to select the optimal attribute according to the historical data set to classify the decision. In random forest, the first step is to randomly select the attribute range for subsequent decision tree training according to the given subset attribute number from the training data set.

If it is specified that the number of attributes involved in the training data set is \( d \), and the set number of subset attributes is \( k \). The random forest algorithm degrades into a decision tree algorithm, if \( d = k \). The current study shows that when the value \( d \) is \( \log_2 d \), the optimal generalization error partition result can be obtained. Therefore, the number of attributes of subset is generally set as \( \log_2 d \).

(3) decision tree training
Based on the given number of attributes of a subset, the attributes of each decision tree are selected first and trained according to the traditional decision tree algorithm. Considering that decision tree training is a relatively mature technical method, its implementation process will not be described here, more details could be found in ref. [10].

(4) summary of decision results
This step refers to the process of summarizing the decision results of each decision tree in the random forest agent in the practical application process to form the overall decision results. The main summary methods that can be adopted include the average method, the comprehensive method and etc. It can be adjusted according to actual needs.

3. Random forest assessment method for power grid operation risk
In this section, random forest algorithm would be applied to power system real-time operation risk assessment problem. Combining with the actual operation risk analysis, a real-time operation risk based on random forest decision implementation framework is constructed.

As shown in figure 2, the key point of traditional power grid operation risk analysis is to conduct panoramic model of power grid operation based on the huge power grid operation data and other relevant data collected, scan typical fault types, evaluate the possible hazards and consequences of different faults and calculate operation risk [11-12].

![Diagram](image.png)

**Figure 2.** Traditional power grid operation risk analysis process

Therefore, this paper will use the historical data of risk assessment to train and generate random forest agents, study the internal relationship between basic data and operation risk assessment results and realize efficient decision analysis from basic data to risk results. As shown in figure 3, there are three implementation elements in this method.

(1) power grid operation data processing
The historical data of power grid operation are used to form the historical data set of power grid operation risk according to the judgment analysis method shown in figure 2. Through data processing, the training set and verification set are generated. It should be noted that the data samples of the
training set and verification set should not intersect. Similarly, real-time data processing for grid operation generates test sets. 

(2) random forest training and improvement

Given the number of decision trees in the initial random forest, the random forest generation training method introduced in figure 1 is used to generate random forest agents for power grid operation risk determination. Using the agent, the operational risk of each sample in the verification set was determined, and the accuracy of the verification set was determined by comparing with the calculated risk value in offline mode. If the accuracy of the verification set meets the requirements, the output results; Otherwise, increase the number of decision trees and retrain to generate random forest agents until the accuracy meets the requirements.

(3) real-time running judgment analysis

The real-time data of power grid operation can be analysed by the random forest agent, to obtain the real-time risk assessment results of power grid operation.

\[\text{Figure 3. Power grid operation risk analysis process based on random forest}\]

4. System development and implementation benefits

4.1. System development

A real-time operational risk assessment system (RTORAS) has been built in Guizhou power grid of China and the method proposed in this paper has been applied in practice. As shown in the figure 4, the RTORAS is deployed in security operation zone III. There are four related systems that support RTORAS and provide the basic data.

(1) Energy management system (EMS), which is deployed in security operation zone I, mainly provides power system real-time operation data, including equipment real-time operation state, real-time load and power plant generation output;

(2) Forecast management system, which is deployed in security operation zone II, mainly provides ultra-short-term load forecast and new energy power forecast data;

(3) Meteorological information system, which is deployed in security operation zone III, is responsible for providing meteorological information, including precipitation, wind and other data;

(4) Operation management system (OMS), deployed in security operation zone III, is responsible for providing basic data, including generation unit list, power transmission and transformation equipment list and so on.

\[\text{Figure 4. System framework}\]
4.2. Operation mechanism and implementation benefits
This system is mainly used for real-time power grid operation state scanning. Compared with the traditional risk assessment analysis, the real-time power grid operation risk assessment method based on random forest algorithm takes less assessment time, with less than 5 seconds each assessment time. The traditional power grid analysis method is only used when an unexpected power system operation risk status has been detected by the assessment method based on random forest algorithm. Since January 2018, the RTORAS has been put into practice for nearly two years. During the period, a total of 15 abnormal operation risks has been found. In these events, the comparison between the risk assessment results of the proposed method and those of the traditional method is shown in figure 5. It can be seen that the deviation of the evaluation results of the proposed method is less than 5% compared with the traditional evaluation results, while the calculation time is only within 5 seconds. Therefore, the proposed method and the traditional evaluation method have good complementary characteristics. The combination of the two can further improve the efficiency of real-time risk assessment.

![Figure 5. Comparison of assessment results](image)

5. Conclusion
Compared with the traditional operation evaluation method, the real-time operation risk evaluation method based on random forest proposed in this paper has the advantages of higher computing efficiency and lower computing cost. This method is suitable for real-time rolling monitoring of power grid operation and can be used as a reference for dispatching operators. On this basis, a more comprehensive assessment method is further applied to accurately assess the risk source and risk level when a greater hidden risk is found.

Reference:
[1] Feng Y, Wu W, Zhang B and Li W 2008 IEEE Transactions on Power Systems Power System Operation Risk Assessment Using Credibility Theory 23 1309-1318
[2] Negnevitsky M, Nguyen D H and Piekutowski M 2015 IEEE Transactions on Power Systems Risk Assessment for Power System Operation Planning With High Wind Power Penetration 30 1359-1368
[3] Wang C, Liu F, Wang J, Wei W and Mei S 2016 IEEE Transactions on Sustainable Energy Risk-Based Admissibility Assessment of Wind Generation Integrated into a Bulk Power System 7 325-336
[4] Ansari A and Chung C Y 2019 IEEE Transactions on Power Systems A Hybrid Framework for Short-Term Risk Assessment of Wind-Integrated Composite Power Systems 34 2334-2344
[5] Li X and Jiang C 2011 IEEE Transactions on Power Systems Short-Term Operation Model and Risk Management for Wind Power Penetrated System in Electricity Market 26 932-939
[6] Preece R and Milanović J V 2015 IEEE Transactions on Power Delivery Risk-Based Small-Disturbance Security Assessment of Power Systems 30 590-598
[7] Xu Y, et al 2016 IEEE Transactions on Neural Networks and Learning Systems Assessing Short-Term Voltage Stability of Electric Power Systems by a Hierarchical Intelligent System 27 1686-1696
[8] Liu X, Shahidehpour M, Cao Y, Li Z and Tian W 2015 IEEE Transactions on Smart Grid Risk Assessment in Extreme Events Considering the Reliability of Protection Systems 6 1073-1081
[9] Li H, et al 2019 IEEE Access An Integrated Online Dynamic Security Assessment System for Improved
Situational Awareness and Economic Operation 7 162571-162582

[10] LI Guanghua, LI Junqing, ZHANG Liang, et al 2019 Computer Science Feature Selection Method Based on Ant Colony Optimization and Random Forest 46 212-215

[11] Guangbin Lian and R. Billinton 1994 IEEE Transactions on Power Systems Operating reserve risk assessment in composite power systems 9 1270-1276

[12] Dawn S, Kumar M, Tiwari A, Kumar Goswami and R. Panda 2019 IEEE Transactions on Sustainable Energy An Approach for System Risk Assessment and Mitigation by Optimal Operation of Wind Farm and FACTS Devices in a Centralized Competitive Power Market 10 1054-1065