Research on Power System Performance Evaluation Based on Machine Learning Technology

Qijiu Yang
Tourism and Culture College of Yunnan University Faculty of Information Lijiang, Yunnan, 674199, China
yqj8799@163.com

Abstract. Power systems often suffer from various large disturbances during operation, especially grounding and short-circuit faults of operating lines, which may lead to transient instability of the system. In view of the fact that the existing relay protection is difficult to be fully applied to the power system with high permeability distributed energy, the machine learning algorithm is applied to the relay protection of the power system. Enhance the robustness of the model to noise; In the training, more weight is given to the unstable samples to balance the influence caused by the difference in the number of samples. In addition, a regular term is introduced into the loss function to control the complexity of the model and reduce over-fitting, thus adapting to various operating conditions of the power system. By comparing the difference between measured data and estimated data to detect bad data, the machine learning method is more intelligent than the traditional method. The research results show that the transient stability evaluation method based on incremental learning of support vector machine greatly reduces the learning time while maintaining the evaluation performance, and is a promising online learning algorithm for transient stability evaluation.

1. Introduction
At present, energy conservation and emission reduction, green energy and sustainable development have become the focus of attention of various countries. The contradiction between the increasing energy demand and the pressure of global resources and environment is deepening. With the deepening of the electricity market, users' requirements on the reliability and quality of electricity are also continuously improving [1]. Therefore, it is an important link to ensure the safe and stable operation of the power system to make fast and accurate judgment on the transient stability of the power system and provide basis for formulating reasonable dispatching control strategies. Big data provides a brand-new development platform and unprecedented opportunities for the power communication industry. However, big data processing undoubtedly brings new challenges to communication for the increasingly information-based electricity [2]. The data transmission, data acquisition and analog-to-digital conversion involved in this process are all likely to produce many errors. Not only that, each process may sometimes be interfered or malfunctioned. Therefore, due to the limitation of computer memory capacity and computing speed, it is only suitable for the case of a small number of samples, and cannot handle the problem of a large amount of data at all [3]. Usually, the purpose of transient stability analysis is to check the stability of the system under the specified operation mode and failure mode, and to put
forward corresponding requirements for relay protection, automatic devices and various measures. In view of the fact that the traditional protection algorithm is difficult to adapt to the development of power system, this paper applies machine learning algorithm to the performance evaluation of power system.

2. Methodology

As we all know, the progress of machine learning and data mining technology is of great significance to computer science and even the whole field of science and technology [4]. The traditional main power grid has a relatively fixed operation mode, and the size and direction of power flow are easier to determine, which is conducive to the realization of relay protection in the power system. In addition, the actual power system data obtained through the wide area measurement system often contain a certain level of noise, which will bring the possibility of over-fitting to the model, thus reducing the generalization ability of the model. In this paper, each site is taken as a node. If there is a connection between the two sites, a line is connected between the two sites as the edge of the network, so as to establish a network model based on adjacent nodes. To some extent, because there is no similar concept that can be extended as a source concept, inductive learning has a larger amount of reasoning than analog learning [5]. According to the limited sample information, it seeks the best compromise between the complexity of the model and learning ability in order to obtain the best promotion ability. Especially in small sample wood, high dimension and nonlinear data space, support vector machine has better generalization ability [6]. In addition, the fault current injection capability of the distributed power source depends on the equivalent impedance of Davinan after it is incorporated into the distribution network, while the capacity of the distributed power source determines the equivalent impedance of its access. Noise will bring over-fitting, which will further affect the model's learning of small class sample features. Therefore, it is not comprehensive to consider only one of the problems, which will lead to the lack of practical application value of the evaluation model.

A general model of machine learning is shown in Figure 1, wherein the generator independently generates a random vector \( x \in \mathbb{R}^n \) from a fixed but unknown probability distribution function \( F(x) \); The trainer returns an output value \( y \) for each vector \( x \), and the output is generated based on the same fixed but unknown conditional distribution function \( F(y | x) \); The learning machine can realize a certain function set \( f(x, a) \) \( a \in \Lambda \), where \( \Lambda \) is the parameter set [7].

![Figure 1. According to the general model of sample wood learning](image)

In short, learning is to select the function that best approximates the response of the trainer from the given function set \( f(x, a) \). This requires measuring the loss \( L(y, f(x, a)) \) between the actual response \( y \) given input \( x \) and the response \( f(x, a) \) given by the learning machine, and its mathematical expectation is [8]:

\[
R(a) = \int L(y, f(x, a))dF(x, y)
\]  

(1)

The definition of learning problem is as follows: there is a probability measure \( F(z) \) defined on the space \( z \), and the set \( Q(z, a), a \in \Lambda \) of functions are considered. The goal of learning is to minimize the risk functional:
The probability measure $F(z)$ is unknown, but a certain number of independent identically distributed trees are given.

Let the output $y$ of the trainer take two values $y = \{0, 1\}$, and let $f(x, a) \in \Lambda$ be the indicator function set. Corresponding loss function:

$$L(y, f(x, a)) = \begin{cases} 0 & \text{if } y = (x, a) \\ 1 & \text{if } y \neq (x, a) \end{cases}$$

(3)

Let the output $y$ of the trainer be a real value, and let $f(x, a) \in \Lambda$ be a set of real functions, including a regression function that minimizes the risk function under the loss function $L(y, f(x, a)) = (y - f(x, a))^2$:

$$f(x, a_0) = \int ydF(y | x)$$

(4)

At present, there are some basic parameters such as degree, shortest path, and betweenness and so on in characterizing the statistical properties of complex network structures. These parameters show the characteristics of the network from different levels. It is a method of generating abstract concepts from concrete examples. Example learning is the main research object of machine learning and the most basic learning ability of human beings. The early time series method and regression analysis method have small calculation amount and high speed, but the model is too simple to simulate complicated and changeable power load [9]. The research on transient stability assessment is limited to the methods of "machine learning to summarize specific problem models from limited observations" and "data analysis to discover various relationships implied in data from limited observations". Different locations of distributed power supply will also affect the magnitude and direction of fault current. Therefore, the impact of distributed power supply access mainly depends on the size, type and access location of its power supply capacity. It can transfer the knowledge learned in one scene to another scene for application, which means that the model trained based on the original topology structure can have good generalization performance in the system after the topology structure is changed. The betweenness reflects the role and influence of the node or edge in the whole network. The greater the betweenness, the greater the role and influence of the node or edge in the network [10]. For each subset, the minimum empirical risk can be found. When considering the empirical risk and the confidence range between subsets, compromise consideration should be made between subsets to obtain the minimum actual risk, which is the idea of structural risk minimization.

As mentioned earlier, judging the transient stability of the system belongs to the problem of pattern recognition. The machine learning model for constructing transient stability evaluation is shown in Figure 2, which specifically includes:

Figure 2. Machine learning model for transient stability assessment.
In order to improve the performance of power system transient stability assessment models, combining multiple or multiple assessment models is an effective way to improve classification accuracy by aggregating the predictions of multiple assessment models. Impact on upstream line protection. The fault current of the upstream line protection device flowing through the fault point is unchanged, and the fault is removed through normal operation of protection; For the problem that it takes a lot of time to generate data sets, on the one hand, if the generalization problem of the model is solved, there is no need to repeatedly generate different data sets. On the other hand, cloud computing and parallel computing technologies are used to simulate power system in time domain. The reliability of the power communication network is affected by the devices in the network, personnel management and other factors, and plays an important role in the load-bearing traffic of the power communication network. Power system automation devices are mainly composed of discrete analog transistor circuits. These analog automation devices play an important role in the automation level of power system and ensure the normal and safe operation of power system. However, the final solution of this kind of method is too dependent on the initial value, its convergence speed is slow, and the number of hidden nodes in the network is difficult to determine. According to the sampling distribution of training samples, samples are extracted to obtain a new sample set. Then, a classifier is induced from the training set and used to classify all samples in the original data set. If a certain data has been accurately classified, the probability of being selected is reduced in constructing the next training set. Through the analysis of the main technical indexes mentioned above, it can be seen that when the single-phase short circuit occurs, a reasonable setting of neutral grounding impedance value can reduce the impact of distributed power supply access to the original distribution network protection.

3. Result Analysis and Discussion

The main power grid with distributed energy has changed the power distribution and impedance distribution of the traditional power grid. The classical relay protection method is difficult to fully adapt to the relay protection with distributed energy. Under normal circumstances, the connection weights between nodes are iterated through a specific optimization algorithm. The iteration of the network is often terminated when reaching a certain training precision or a certain number of iterations. Due to the high computing power of microprocessor, the new algorithm can be applied in power system, which plays an important role in improving the accuracy of power system measurement system, the reliability of control system and the level of intelligence. On the contrary, if a certain data is not correctly classified, its weight will be increased. Machine learning, as a multidisciplinary interdisciplinary subject, involves many disciplines such as probability theory, statistics, approximation theory, convex analysis and algorithm complexity theory. The root node contains all the training samples, and each child node contains samples divided according to different characteristics. The most critical step in decision tree learning is how to divide features. When the extreme learning machine algorithm is used to predict the time series, the learning speed is accelerated to a great extent, and the prediction accuracy becomes higher and higher. At the same time, compared with transistor automatic devices, microprocessor automatic devices expand new functions, especially the self-diagnosis capability of automatic devices, which plays an important role in improving the service life of automatic devices and their own reliability.

Xn is the input feature vector. The importance of the component classifier ci depends on its error rate \( \varepsilon_i \), which is defined as:

\[
\varepsilon_i = \frac{1}{N} \sum_{j=1}^{N} w_j I(C_i(x_j) \neq y_j)
\]

(5)

Among them, \( w_j \) is the weight of sample wood points, and I (.) is indicator function. The importance of the component classifier Ci is given by the following parameters:
The parameter $a_i$ is also used to update the weight of the training sample wood. The update mechanism is as follows:

$$w_i^{(j+1)} = \frac{W_i^{(j)}}{Z_j} \times \begin{cases} e^{-a} \text{ if } C_j(x_i) = y_j \\ e^{a} \text{ if } C_j(x_i) \neq y_j \end{cases}$$  \hspace{1cm} (7)$$

The final overall classification decision is obtained by weighted average of each component classifier $C_i$:

$$g(x) = \text{sign} \left[ \sum_{j=1}^{J} a_j C_j(x) \right]$$ \hspace{1cm} (8)$$

A great deal of transient stability analysis is required for power system planning, design and operation. Traditional stability analysis is carried out under the condition that the component parameters, operating conditions and interference modes of the system have been given. The deterministic stability analysis machine learning relies on excellent data sources and needs good data support. At present, the rapid development of power big data and super-powerful computers and other technologies has improved a greater number of better training methods for machine learning. Due to the scarce data of unstable operation samples in actual power systems, conventional classifiers will learn more about the characteristics of stable samples in the training process, resulting in insufficient identification ability of unstable samples. However, the extreme learning machine algorithm is not devoid of any shortcomings. In the case of solving the same problem, the different selection of activation functions will lead to large differences in the final results and cannot obtain accurate prediction values. Therefore, the extreme learning machine algorithm also has certain limitations. To combine. Due to the lack of unified standards for substation automation systems, different communication protocols of equipment from different manufacturers, and complicated communication control, it is difficult to effectively realize high-speed, reliable and accurate information exchange between intelligent electronic devices (IEDs), which requires adding new software and hardware to connect IEDs, thus weakening the advantages of substation automation to a certain extent. Therefore, probabilistic transient stability assessment is a breakthrough from certainty to probability in power system stability analysis, an important supplement to traditional certainty analysis, and has gradually attracted the attention of researchers in various countries.

In order to compare the effectiveness of the selected feature subset, the original feature set is reduced to 12-dimensional feature subset $b_2$, which takes 95% of the variance of the original data set. An arbitrarily selected feature subset $b_3$. Node system recommended feature $b_4$, model training is the same as 39 bus system, and test results are shown in Table 1 and Figure 3.

| Input feature (dimension) | Test time of single sample (ms) | Test accuracy (%) |
|---------------------------|---------------------------------|-------------------|
| Original feature B (25 dimensions) | 1.12 | 93.25 |
| Feature subset B1 (10 dimensions) | 0.94 | 94.15 |
| PCA subset B2 (12D) | 1.24 | 92.11 |

Table 1. Test system results
Because the analytical method adopts strict mathematical means, the calculation results are highly reliable. The disadvantage of the analytical method is that when the system scale increases, the amount of computation increases exponentially with the increase of the system scale. On the one hand, the stored power big data are stored in the knowledge database, and the functions of knowledge database inquiry, recording and diagnosis are provided. At the same time, the complicated operation of the power system will increase the diversity of samples, which will lead to over-fitting of the model. On the one hand, the reliability of the power-to-communication network is affected by the devices and personnel management in the network, and on the other hand, it plays an important role in carrying the traffic volume of the power-to-communication network. Therefore, the research on the reliability of network topology needs a comprehensive, complete and reasonable analysis from the two aspects of network devices and traffic load. The function and data object of substation automation system are modeled by object-oriented technology, and abstract communication service interface independent of network structure is adopted, so that equipment and data object can be described by themselves. The basic idea is to first establish a probability model or a random process so that its parameters are the required solution of the problem, then calculate the statistical characteristics of the required parameters through observation or sampling tests of the model or process, and finally give the approximate value of the solution. The system improves the timeliness and accuracy of on-site exception handling, eliminates the chance of data, refines the inevitability of data, and provides accurate support for identification and judgment of various complex power system relay protection states under distributed energy sources.

4. Summary
With the high penetration of distributed energy into the main power grid, the operation of the power system becomes more complicated and the relay protection is more difficult. This paper applies machine learning method to the performance evaluation of the power system. Each equipment and line of power communication network are abstracted into nodes and edges, a network topology model is established, and various parameters such as average path length, clustering coefficient, betweenness, degree and degree distribution are calculated. Then the reliability of network topology structure is analyzed, and the selection principle and measurement index of reliability index system are given. It lays a foundation for condition-based maintenance of equipment, and has certain positive significance for improving the reliability detection of equipment and ensuring the safety, stability and economy of power system operation. At the same time, compared with traditional equipment detection methods, the software takes less time. When the operation mode and topology structure of the power system change, the previously
trained model may no longer be available. Therefore, how to make use of transfer learning and other technologies to make the model have inheritance and wider adaptability is the focus of future research.

References
[1] Jang D W, Lee S, Park J W, et al. Failure Detection Technique under Random Fatigue Loading by Machine Learning and Dual Sensing on Symmetric Structure [J]. International Journal of Fatigue, 2018, 114.
[2] Thai D K, Tu T M, Bui T Q, et al. Gradient tree boosting machine learning on predicting the failure modes of the RC panels under impact loads [J]. Engineering With Computers, 2019 (2): 1 - 12.
[3] Nateghi R. Multi-Dimensional Infrastructure Resilience Modeling: An Application to Hurricane-Prone Electric Power Distribution Systems [J]. IEEE Access, 2018, PP (99): 1 - 1.
[4] Wolfgang K. Artificial Intelligence and Machine Learning: Pushing New Boundaries in Hearing Technology[J]. Hearing Journal, 2019, 72(3):26.
[5] Kalvoy H, Tronstad C, Ullensvang K, et al. Detection of needle to nerve contact based on electric bioimpedance and machine learning methods.[J]. Conf Proc IEEE Eng Med Biol Soc, 2017: 9 - 12.
[6] Yuwei S, Zhao M A, Chenyang P, et al. Study of a Novel Machine Learning Method Embedding Expertise Part 1: Proposals and Fundamentals of Guiding Learning[J]. Proceedings of the CSEE, 2017, 37 (19): 5560 - 5571.
[7] Karam R, Paul S, Puri R, et al. Memory-Centric Reconfigurable Accelerator for Classification and Machine Learning Applications [J]. ACM Journal on Emerging Technologies in Computing Systems, 2017, 13 (3): 1 - 24.
[8] Xi Q, Rauschenbach T, Daoliang. Review of Underwater Machine Vision Technology and Its Applications [J]. Marine Technology Society Journal, 2017, 51 (1): 75 - 97.
[9] Cao H, Cai J. Distributed Multi-User Computation Offloading for Cloudlet based Mobile Cloud Computing: A Game-Theoretic Machine Learning Approach [J]. IEEE Transactions on Vehicular Technology, 2017, PP (99): 1 - 1.
[10] Rocha H R O, Silvestre L J, Celeste W C, et al. Forecast of Distributed Electrical Generation System Capacity Based on Seasonal Micro Generators using ELM and PSO [J]. IEEE Latin America Transactions, 2018, 16 (4): 1136 - 1141.