1. Introduction

It is necessary to determine the probability distribution of outage scale for cascading failure analysis and risk assessment. It is always an important part of traditional power system reliability evaluation to calculate the historical fault statistics. However, statistical computation requires a large amount of sample data, especially the calculation of the probability distribution of outage scale, and it must be used for many years, which limits its application in annual risk management. There are many differences in topology, power grid scale, installed capacity, intrinsic connection degree, load level and other technical characteristics of different power grids. Similarly, the planning and operation of power systems or power markets among different grids are different, especially between countries and countries, and the operating mechanisms are different. Fault record data from different sources, reflects the adequacy and security of different systems, the analysis shows that the reliability characteristics of the power grid, which is different, and display the fault probability distribution of similar scale. Different power grids, the applicable probability methods may vary.

The typical development process of the power system cascading failure is: the initial fault cause and trigger system flow transfer, resulting in some heavy line overload, overload protection and corresponding circuit breaker have been tripping, which evolved into the global fault and local fault. Some lines in the process of fault propagation in resection of little overall impact on the system, not enough to cause blackouts; and some key lines removed, will affect the connectivity of the network, the flow transfer is relatively concentrated, accelerate the development of cascading failures. In addition, the wiring of the actual power system is very complex. If each line of the system is taken into consideration when searching for cascading failure modes, huge amounts of computation can be generated and it is difficult to implement online applications. Thus, identifying the key lines in each
link of cascading failures can quickly realize the risk assessment of cascading failures in power systems.

2. Basic Principles of Type Identification and Classification of Cascading Failures

2.1 Basic Principles of Data Mining

The word “data mining” first came from statistics, suggesting blindness in action. Before the data mining technology into the industrial equipment condition monitoring and fault diagnosis, first from the angle of fault diagnosis on the four main stages of data mining process model understanding and description, so as to avoid the blindness of the application. Data mining is best suited for applications where data is changing at a weak level and ordinary human experts cannot. For example, in equipment condition monitoring and fault diagnosis of equipment, the operator may have a good understanding of the overall characteristics of equipment system, but the actual monitoring data changes knowledge update equipment parts and processing objects caused by it is difficult to obtain. This phase is generally divided into four parts, data selection, data cleaning, data transformation and data reduction. Because of the characteristics of the data monitoring system usually has diversity, complexity and mass, therefore must be based on the mining target, choose the corresponding data set in the monitoring database and history database, cleaning, transformation and reduction. Fault diagnosis is one of the most important, critical and difficult problems in equipment fault diagnosis. In order to fundamentally solve the key problem of feature extraction, people usually get more information from the depth processing of signals by means of signal processing theory, methods and techniques.

2.2 Basic Principles of Dispatching and Operation

According to the characteristics of distribution network repair command service and the construction experience of distribution network dispatching informatization, the following key construction modes of distribution network fault diagnosis system are as follows. Independent mode refers to a model that is independent of other systems. The advantage of this system is that it can deploy independently without interference from other systems. Its disadvantage is that the investment is large, the interfaces needed to be developed are large and the workload is large, so it is difficult to integrate and integrate the distribution network scheduling and the daily management system. Based on the distribution management system construction model, it is to expand the distribution network fault analysis function. This model has access to distribution network real-time data and other advantages, but there are great difficulties to exchange data with the area of the geographic information system, marketing management system, the interface is numerous, basic data, repair command data and dispatching management system of distribution network cannot be tightly integrated. Although the analysis module is deployed in the automation system of distribution network analysis on mirror base fault in distribution network, can effectively solve the problem of data exchange system, but is equivalent to a set of multiple independent systems, distribution automation system independence will be reduced. In addition, due to the high cost of distribution automation system construction, only a few distribution networks are equipped with distribution automation system. Therefore, this model cannot be extended for a while. The construction model of dispatching operation management system is to enlarge the function of fault diagnosis of distribution network. This model not only can share the basic data. The interface can avoid duplication of development, reduce the data maintenance, but also can inherit the technology system of smart grid dispatching the control system and the results of direct coordination to achieve a variety of business of distribution network, and be able to distribution network scheduling, distribution network planning, power outage operation and distribution network repair business with full command. In addition, the model is more convenient and low cost, which does not affect the security of other systems.
2.3 Basic Types of Cascading Failures

The hidden fault of protective device, also called hidden fault, is the main factor to trigger cascading failures. In the normal operation of power grid, hidden faults are not easily detected, but can be triggered when a fault or other abnormal state occurs, leading to protection device errors or inappropriate removal of system components. The branch fault causes the power flow of the network to flow. When the current flowing through the branch exceeds a certain limit, the hidden fault may be triggered, and the corresponding branch is cut off. When the above process occurs in many stages and the harm reaches a certain extent, it may lead to a large area blackout. The different causes of cascading failure process of different lengths, but the accident process is a series of elements have been disconnected, the physical process can be simplified into one after another slowly breaking, breaking the rapid succession and restoration. The main risk assessment for the protection of hidden failures so that the slow development stage in the process of failure have been opened for the initial accident and hidden failures lead to incorrect action of protection; at the same time that the rapid development of the process has been opened broken period lead to lead to action and protection hidden failure load protection incorrect action for the trend of large-scale transfer. The blackout accident evolution analysis, found in the fast-continual outage stage, as the development process is very rapid. The stage will have sufficient time to adjust to the system.

3. Type Identification and Risk Evaluation of Cascading Failures

3.1 Type Identification of Cascading Failures

When a fault occurs, the amplitude, probability distribution, frequency component, the main energy spectrum peak position of the signal, etc., will change in time and frequency domain. Therefore, the fault information of biological shock can also be described by the signal waveform in the time domain and the size and distribution of the spectrum energy in the frequency domain. In this paper, 1 time domain characteristic parameters and frequency domain characteristic parameters are extracted from the total leakage current signals before and after an electric shock. The feature vectors which are composed of them are of high dimensionality, and some of the characteristic parameters are related to each other, and redundant feature parameters will exist. Therefore, in order to avoid the curse of dimensionality to improve the performance of classification, it is necessary to eliminate redundant feature parameters before inputting an electric shock fault type identification, then input the classifier for recognition. The factor analysis as an effective analysis tool in multivariate statistics, the basic idea is to use a variable representation of a class of variable close contact in order to find a small number of variables are not relevant to reflect most of the information of the original variables, this eliminates the multicollinearity, achieves the effect of dimensionality reduction. However, the prediction of sequential movement of automatic devices is still difficult. The most convincing scheme, of course, is to simulate the dynamics of the system after the previous break, then compare the action values of various automatic devices, and then determine whether the automatic device is moving. However, there are still a lot of patterns that can be used, and the presence of uncertainty and the search for more patterns may be considered less acceptable in larger systems.

Risk warning, risk control function unit power dispatching and maintenance plan based on offline mode, and emergency supplies and other emergency plan arrangement, outage management functions are required to fully consider the conflict between the two operating costs and security risks, and maintain the level of risk within a tolerable range, operation cost and safety the risk of compromise: the multi-objective optimization method on the problem of modeling, decision makers to consider the relationship between risk attitude and incentive effect and restraint effect, policy-makers to maximize the utility function as the criterion, from a series of. The optimal operation strategy is selected in the optimal solution to obtain the expected generation cost and operational risk level. Through the risk sensitivity analysis, obtains the risk sensitivity matrix the control variables and the dynamic relationship between all kinds of security risk index, in order to improve the calculation efficiency of risk decision model of online mode.
A more feasible approach is to screen the results based on the comparison of values and probability factors. The principle is that the total number of consecutive searches is within the allowable range of the computational resource or within the fault scanning time set by the security defense system. We use the residual current action protection device designed by the task group, and the electric shock physics test system platform. The total leakage current, the period before the electric shock and the signal waveform after the shock are shown in the figures.

Figure 1 Total Leakage Current Signal of the First Type of Cascading Failures

Figure 2 Total Leakage Current Signal of the Second Type of Cascading Failures

Figure 3 Total Leakage Current Signal of the Third Type of Cascading Failures

4. Occurrence Probability of Cascading Failures
As the power system is a complex system running in real time, the probability of its component failure is influenced by many factors, so it is difficult to put forward an accurate mathematical model. For the sake of simplicity, this article considers only the following 2 main factors. Power flow variation in the component after overload and forward fault. Second of these factors are the main cause of hidden trouble in relay protection. The first I reflects and quantifies the assessment. As for the risk assessment of cascading failures, the formula for the probability of occurrence of many faults is as follows:

\[ p_i(x_1, x_2) = \frac{1}{1 + e^{14.09 - 6.98x_1 - 11.26x_2}} \]

\[ x_1 = \frac{P_i}{P_{imax}} \]
Assume \( E = \{E_1, E_2, ..., E_n\} \) is a set of the cascading failures, the occurrence probability of these cascading failures is:

\[
P(E) = P(E_1)P(E_2) ... P(E_n)
\]

Risk indicators focus on energy shortfall, disaster level and influence range of distribution, static security risk of system level and regional level, the analysis results can be used as a basis for hours and minutes in grid security risk management. Because a system can be obtained in advance of the recent safety risk trend, is conducive to the dispatch center professional department staff and leadership focus on the future network security risk in one week, and the high risk of incident contingency plans well in advance and special exercises work. Once, two times the running status of online tracking system, meteorological disasters, wind power and load uncertainties and changes of power grid operation mode, the risk identification and evaluation of the content of dynamic adjustment and refinement. To determine the sub control area and set off the risk status, find out the weak links of power grid, risk warning information, and according to the warning level automatically and recursively on the current security issues of potential grid optimal control decision. The coordination of operational risk management at different time scales is a typical risk decision problem.

If on the stage of the control work is not sufficient, the next stage will increase the difficulty of control measures on the stage. This paper argues that the R before the operation risk management must take into account the economic indicators of the system, focus on the overall situation in the next few roughly control grid operation risk, the risk control and risk transfer from unknown to hours ago and minutes ago; risk control is an hour ago to run the risk index continuously the rolling optimization control of the process, should be a reasonable choice of control scenario set and control measures to avoid the small probability and high risk events, and to realize the optimization and coordination of economic risk and risk control; the last line of defense is minutes before the risk control system, its purpose is to the current security issues have been exposed and the forecasting time potentially high probability of high risk events of emergency control, maintain the stable operation of the system, reduce the maximum working pressure of the dispatcher at run time .

Security risk assessment should be from the angles of power grid, equipment, personnel, management and environment of uncertainty facing the comprehensive consideration of the system, the dynamic generation of contingency set, through the static security analysis and dynamic security analysis and calculation of the severity, combined with risk rules, get all kinds of risk index. There are many uncertainties in the operation of power plants, and these uncertainties are often expressed as random, fuzzy and multiple uncertainties. Because the average rate constant cannot describe fault prediction time uncertain factors for the forced outage, using probability theory and mathematical statistics, uncertainty theory and calculation method for characterization of these factors. It is a difficult and key technology to establish a switching time change model with consideration of micro meteorology, disaster and equipment health. In order to improve the computational efficiency, the expected fault set and its probability are dynamically generated according to the time scale of the risk assessment and the corresponding system state selection method.

5. Severity Degree of Cascading Failures

The severity measure is used to describe the consequences of cascading failures. The common measures include total load loss, direct and indirect economic loss, and the relative degree of electric quantity. The direct and indirect economic losses caused by cascading failures shall be the losses suffered by all the affected entities. Among them, the loss of equipment caused by the accident, product waste and other explicit losses are easy to estimate, while the social, environmental, ecological and other hidden losses caused by difficult to effectively estimate. After cascading failures, the grid frequency, node voltage and branch current are seriously restricted. The traditional method uses some function mapping of the degree of each electrical quantity to map the global severity measure of power
grid, but the function mapping is difficult to judge accurately and is affected by subjective factors. Load is one of the most important variables in power grid. When a line in the system due to maintenance, failure is out of operation, will cause the network flow transfer; but not every line in the trend will change greatly, the trend of changes in some lines in the very small or almost unchanged. If we can identify the trend of greater transmission section lines, can narrow the scope of network security analysis, security analysis to accelerate the speed, so there is more time for decision making. When a disconnection overload branch system, parallel flow with the same power often will flock to the area or the load zone, the key to identify the flow increase with the larger line is parallel section find overload lines, based on the trend of parallel lines in increasing the amount of calculation section.

When the electric quantity exceeds the limit, the measures adopted are mostly related to load, and the amount of load loss is easy to count. The measure of a cascading failure severity function should follow the following principles: quality can reflect the security situation of the system itself; whether it can reflect the economic consequences of cascading failure; whether you can get to calculate the required data conveniently; the realization of online application. In this paper, we choose the indexes of low voltage severity, over load severity and isolated load to establish cascading fault severity function.

The low voltage severity index reflects the degree of deviation from the rated value of each busbar in the system after the occurrence of the fault:

$$\text{Sev}(V_j) = \begin{cases} 
0.95 - V_j & V_j \leq 0.95 \\
0.95 & V_j > 0.95
\end{cases}$$

When there are many overload elements in the power system, the system is in an extremely unstable state, and a little disturbance may lead to blackouts. Defines the overload severity function as follows:

$$\text{Sev}(P_k) = \begin{cases} 
P_k - 1.0 & P_k \geq 1.0 \\
0 & P_k < 1.0
\end{cases}$$

The cascading trip of power system can cause the loss of power supply in some power systems, and the system loses a great deal of load, resulting in huge economic losses. The isolated load severity function is:

$$\text{Sev}(P_{L}) = \frac{1}{N} \sum_{i=1}^{N} P_{L}(i)$$

6. Conclusions

The fault identification and risk prediction of distribution network based on data mining can not only effectively support the rapid and active rush repair of distribution network fault, but also make the dispatching and repair command of the distribution network achieve seamless integration. The identification of cascading faults greatly shortens the fault-finding time and improves the efficiency of fault repair of power supply.

Acknowledgement

Fund Project: National Natural Science Foundation of China (71671065); State Grid Corporation of Science and Technology Project(5204BB1600CP).

References

[1] Huang X, Vodenska I, Havlin S, et al. Cascading failures in bi-partite graphs: model for systemic risk propagation[J]. Scientific reports, 2013.

[2] Reynolds-Barredo J M, Newman D E, Carreras B A, et al. The interplay of network structure and dispatch solutions in power grid cascading failures[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2016.
[3] Mukherjee B, Habib M, Dikbiyik F. Network adaptability from disaster disruptions and cascading failures[J]. IEEE Communications Magazine, 2014.

[4] Tan F, Xia Y, Zhang W, et al. Cascading failures of loads in interconnected networks under intentional attack[J]. EPL (Europhysics Letters), 2013.

[5] Koç Y, Warnier M, Van Mieghem P, et al. The impact of the topology on cascading failures in a power grid model[J]. Physica A: Statistical Mechanics and its Applications, 2014.

[6] Wang J W, Rong L L. Robustness of the western United States power grid under edge attack strategies due to cascading failures[J]. Safety science, 2011.

[7] Reynolds-Barredo J M, Newman D E, Carreras B A, et al. The interplay of network structure and dispatch solutions in power grid cascading failures[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2016.

[8] Koç Y, Warnier M, Van Mieghem P, et al. The impact of the topology on cascading failures in a power grid model[J]. Physica A: Statistical Mechanics and its Applications, 2014.

[9] Zhao J, Li D, Sanhedrai H, et al. Spatio-temporal propagation of cascading overload failures in spatially embedded networks[J]. Nature communications, 2016.

[10] Kornbluth Y, Lowinger S, Cwilich G, et al. Cascading failures in networks with proximate dependent nodes[J]. Physical Review E, 2014.

[11] Rahnamay-Naeini M, Hayat M M. Cascading Failures in Interdependent Infrastructures: An Interdependent Markov-Chain Approach[J]. IEEE Transactions on Smart Grid, 2016.

[12] Huang Z, Wang C, Nayak A, et al. Small cluster in cyber physical systems: Network topology, interdependence and cascading failures[J]. IEEE Transactions on Parallel and Distributed Systems, 2015.

[13] Hernandez-Fajardo I, Duela-Osorio L. Probabilistic study of cascading failures in complex interdependent lifeline systems[J]. Reliability Engineering & System Safety, 2013.

[14] Zhang G, Li Z, Zhang B, et al. Understanding the cascading failures in Indian power grids with complex networks theory[J]. Physica A: Statistical Mechanics and its Applications, 2013.

[15] Wang J. Mitigation strategies on scale-free networks against cascading failures[J]. Physica A: Statistical Mechanics and its Applications, 2013.