Key Line Identification of Power Grid Based on Entropy Weight Method of Pearson Coefficient Index Selection

Ming He*, Yanli Zhou
School of Electronic Engineering, Guangxi Normal University, Guilin 541004, China

*Corresponding author e-mail: 303401935@qq.com

Abstract. As one of the important infrastructure of the country, it is very important to identify the key links of the power network and protect them to prevent the blackout. Based on the theory of complex network, this paper proposes a comprehensive multi-attribute identification method for key nodes of power grid. Firstly, the Pearson correlation coefficient is used to select the evaluation index according to the topological characteristics of the power grid, then the entropy weight method is used to get the final decision matrix, and finally adopt the comprehensive method to calculate the order of the important lines in the power grid. In order to compare the advantages and disadvantages of different recognition methods, we compare the evaluation results of electrical method, and further verify the effectiveness of the proposed method in IEEE39 system.

1. Introduction

In recent years, with the continuous expansion of the scale of the power network [1], the transmission capacity and transmission range are also increasing, resulting in the frequent occurrence of large-scale blackouts caused by line failures, which has a serious impact on people's daily life and social activities [2-4]. For a given power grid, its network topology is certain, is the essence of the network, but also reflects the robustness of the network, using complex network theory to analyze the power grid structure and identify vulnerable links to prevent such incidents have a positive role.

At present, more and more scholars use complex network theory to study the vulnerability of network lines, this paper also applies complex network theory to identify the vulnerability of power grid, in order to make up for the incompleteness of using a single index evaluation, this paper uses multi-attribute evaluation method; in the selection of evaluation index combined with the topology characteristics of power grid, select the index with low linear correlation to get the decision-making matrix, then weight each index, and finally get the ranking of network line vulnerability, and through IEEE39 network to analyze and verify the feasibility of the algorithm.

2. Multi-attribute decision-making method based on entropy

2.1. Multiple Attribute Decision Making Description

A typical multiple attribute decision making (MAMD) problem is: a system has different attributes, given a variety of evaluation indexes, each index cannot evaluate a certain attribute in the system to get a decision matrix, and finally get the ranking of each attribute of the system by comprehensive evaluation.
results. For a multi-attribute system, define \( S = (s_1, s_2, s_3 \ldots s_n) \) is the attribute set, defining \( U = (u_1, u_2, u_3 \ldots u_m) \) is the set of evaluation indexes, but the weight of each index is unknown. The evaluation value \( a_{ij} \) of each attributes on evaluating indicators can be obtained, which constitutes the decision matrix \( A = (a_{ij})_{n \times m} \) as follows:

\[
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1m} \\
a_{21} & a_{22} & \cdots & a_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & a_{21} & a_{nm}
\end{bmatrix}
\]  

(1)

After getting the decision matrix, we need to construct the evaluation system and calculate the weight of each index, the commonly used methods are entropy method, analytic hierarchy process, TOPSIS method and so on [5-7]. Entropy method is used to weight the evaluation index in this paper, entropy method is an objective weighting method, in the calculation process according to the evaluation value of each attribute to get the degree of variation of each index, and then get the final weight through information entropy.

2.2. Principle of entropy weight method

Entropy was first used in thermodynamics and was defined as the degree of chaos in the system. In 1948, Shannon used this index to quantify information to obtain the concept of information entropy [8], which is used to describe the uncertainty of information sources. Subsequently, more and more researchers have taken the entropy value calculated in the system as a weight, and have developed extensively in the fields of engineering technology, socioeconomics, planning and construction [9].

In information theory, the probability of the appearance of information is taken as the importance of the information. The greater the probability of occurrence, the less the amount of useful information contained, and the smaller the probability of occurrence, the more important the information is. We set the probability of the occurrence of certain information in the system as \( p_i \) \((i = 1, 2, \ldots, n)\), then the entropy of the system is

\[
e = -\sum_{i=1}^{n} p_i \cdot \ln p_i
\]  

(2)

For a system with \( n \) and attributes and \( m \) evaluation indexes, an evaluation matrix \( R = (r)_{n \times m} \) can be obtained, where the entropy of an index is

\[
e_j = -\sum_{i=1}^{n} p_{ij} \cdot \ln (p_{ij})
\]  

(3)

Where \( p_{ij} = r_{ij}/\sum_{i=1}^{n} r_{ij} \). Then the entropy weight \( \omega_j \) corresponding to this index is

\[
\omega_j = \frac{1-e_j}{\sum_{j=1}^{m}(1-e_j)}
\]  

(4)

The analytical formula (3) can be obtained that the smaller the entropy value \( e_j \) of an indicator, the greater the degree of variation of the values of each attribute under the indicator, the more effective information provided, and the larger the proportion in weights; The larger the entropy value \( e_j \) is, the smaller the value variation of each attribute under the index is, the less effective information is provided, and the smaller the proportion in the weight should be. The entropy weight is obtained through formula (4). We comprehensively evaluate the data obtained from each index to objectively obtain the weight, reduce the influence of subjective factors, and improve the credibility of the empowerment.
2.3. Weighted Index Selection

The principle of entropy weight method was introduced in the previous section, and the results were calculated by the decision index value of each attribute. In a system, the attribute is unique to the system, but the evaluation index is not fixed, if we select some indicators to calculate the value of each attribute is almost exactly the same, the index will lose the value of evaluation, so we need to carry out linear correlation analysis of these indicators when selecting indicators, and select several groups of indicators with small correlation. Here, the indicators are compared using the Pearson Correlation Coefficient (Pearson Correlation Coefficient), which is used to determine the degree of linear correlation between two data vectors, and the determination steps are as follows:

(1) For a given two data vectors and identity matrix: \( X = (s_1, s_2, s_3, ..., s_n) \) \( Y = (s_1, s_2, s_3, ..., s_n) \), \( e = (s_1, s_2, s_3, ..., s_n) \)

(2) The vector is normalized to obtain a normalized matrix, and \( \mu_x \) and \( \sigma_x \) represent the mean and standard deviation of the data in X, respectively

\[
\mu_x = \frac{\sum_{i=1}^{n} x_i}{n}, \quad \sigma_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu_x)^2}{n}}
\]

(5)

In order to facilitate comparison, the vector was standardized, at that time, the standardized matrix of vector \( X \) was defined as: \( \sigma_x \neq 0z_x \)

\[
z_x = \frac{x - \mu_x e}{\sigma_x}
\]

(6)

All normalized vectors have the following characteristics, \( \|z\| = \sqrt{n}, \mu = 0, \sigma = 1 \), and the normalization matrix \( z_y \) of vector \( Y \) is the same.

(3) The Linear correlation coefficient of two vectors \( X, Y \) is defined as:

\[
\rho_{xy} = \frac{z_x^Tz_y}{n}
\]

(7)

Correlation coefficient \( |\rho_{xy}| \in [0, 1] \), the closer \( \rho_{xy} \) to 0, the smaller the correlation, and the closer to 1, the greater the correlation. In summary, the flowchart of identifying key nodes in the power grid is as figure [1]:
3. Identification of key lines in the power grid

In complex network theory, the commonly used evaluation indicators for the network are Degree index, Cohesion Degree index, Closeness Centrality index, Betweenness Centrality index, PageRank index, Subgraph index [9], etc. These indicators reflect the local and overall characteristics of the network. The importance ranking calculated for different structures of the network is different. In order to improve the recognition, we first apply these indicators to the power grid system, and then perform a linear correlation calculation of this indicator in section 1.3 to select the correlation. The smallest indicator. We use the IEEE39 power network to identify key line experiments. The network topology is shown in the following figure [2]. The system has 10 generators and 46 transmission lines. During the identification process, the line of the IEEE 39 system is regarded as a system, and the adjacency matrix $S = (s_{ij})_{46 \times 46}$ of the system line, and the line system is calculated and sorted using the above indicators to obtain a decision matrix $Q = (q_{ij})_{46 \times 6}$, calculate the proportion of the sum of Pearson's correlation coefficient between each indicator $O = (o_j)_{1 \times 6}$:

$$O = [0.1780, 0.1905, 0.1488, 0.1556, 0.1667, 0.1604]$$

![Algorithm flow chart](image)

**Figure 1. Algorithm flow chart**
Figure 2. IEEE 39-Node System wiring diagram.

It can be seen that in the IEEE39 system, the aggregation degree index has the highest similarity with other indicators, so we choose the other 5 indicators as the evaluation indicators of the system. Use the entropy weight method to weight each index, and the weights are

$$\omega = [0.0938, 0.1944, 0.4036, 0.0972, 0.2109]$$

We select the top ten lines of importance ranking for comparison with other recognition methods. The ranking is shown in Table 1. Comparing the method in this paper with the electrical intermediary method [10], considering the voltage rank vulnerability ranking method [11] and the weighted line intermediary method [12], the importance rankings of the lines are compared. There are eight lines in the top ten in the ranking. All in the other three methods, and mentioned in [12] Lines (L15-16, L2-25, L16-17) will cause system instability after a three-phase short-circuit occurs at the high-voltage side. These three lines are included in the identification method of this paper. By comparing electrical methods, it is also verified that Rationality and effectiveness in actual power grids.

| Rank | The method of this paper | Entropy Weight Fragile Value | Electrical betweenness ordering | Voltage level vulnerability ranking | Sorting by the method of medium number of weighted lines |
|------|--------------------------|------------------------------|---------------------------------|-----------------------------------|-------------------------------------------------------|
| 1    | L 15-16                  | 0.256                        | L 16-17                         | L 5-6                             | L 15-16                                               |
| 2    | L 16-17                  | 0.240                        | L 15-16                         | L 15-16                           | L 15-16                                               |
| 3    | L 14-15                  | 0.222                        | L 14-15                         | L 1-2                             | L 16-19                                               |
| 4    | L 4-5                    | 0.212                        | L 16-19                         | L 2-3                             | L 16-21                                               |
| 5    | L 2-3                    | 0.204                        | L 17-27                         | L 4-14                            | L 16-24                                               |
| 6    | L 4-14                   | 0.200                        | L 26-27                         | L 25-26                           | L 17-18                                               |
| 7    | L 3-4                    | 0.196                        | L 2-25                          | L 5-8                             | L 17-27                                               |
| 8    | L 13-14                  | 0.195                        | L 17-18                         | L 2-25                            | L 2-25                                                |
| 9    | L 16-19                  | 0.187                        | L 3-18                          | L 16-17                           | L 2-3                                                 |
| 10   | L 2-25                   | 0.180                        | L 2-3                           | L 21-22                           | L 1-2                                                 |
4. Network efficiency experiment verification

In order to further verify the effectiveness of this method, we use the most commonly used verification method in power network, network efficiency to analyze the results obtained in the previous section. Network efficiency is used to describe the transmission efficiency of the network. In the power system, when a power network is damaged, the primary concern is the transmission efficiency of the final grid. We will compare the network integrity before and after removing the important lines to get the network efficiency of the grid, which is defined as follows:

\[
C_E(v_i) = \frac{E_i}{E}
\]

\[
E = \frac{2}{N(N-1)} \sum_{1 \leq i < j \leq N} \frac{1}{d_{ij}}
\]

(8)

Where \(N\) is the number of nodes in the network, \(d_{ij}\) is the shortest electrical distance between nodes \(v_i\) and \(v_j\) in the network. By comparing the network efficiency values of the remaining networks after removing the lines according to different methods, the more the remaining network efficiency indicates that the damaged nodes are more important to the integrity of the network, and the results are shown in figure [3]. After removing 5 lines, the connectivity of the whole network is half of the original network, and the probability of large-scale cascading failure is also greater, which shows the feasibility of this method.

![Figure 3. IEEE39 system network efficiency.](image)

5. Conclusion

In this paper, based on the idea of multi-attribute decision-making, we use Pearson correlation coefficient method to select the less relevant decision-making indicators, and then use entropy weight method to weight the multi-indicator, considering the contribution of each indicator to the evaluation system. For the incompleteness of the single indicator method used in many similar studies, and the selection of evaluation indicators in the comprehensive multi-indicator method is not considered. Considering the shortcomings of similarity, the evaluation system is improved to make it more perfect and objective, and is tested in IEEE39 power grid system. The comparison results show that the method in this paper is superior to the other three comparison methods. The evaluation model in this paper has
certain guiding significance for improving the reliability of power grid and preventing large-scale cascading.

Acknowledgments
This work was financially supported by the National Science Foundation of P. R. China under Grant Nos.11562003.

References
[1] G. A. Pagani, M. Aiello. From the grid to the smart grid, topologically, J. Physica A: Statistical Mechanics and its Applications, 449 (2016) 160 - 175.
[2] E. Bompart, R. Napoli, F. Xue. Analysis of structural vulnerabilities in power transmission grids, J. International Journal of Critical Infrastructure Protection, 2 (2009) 5 - 12.
[3] N. S. Bhangu, R. Singh, G.L. Pahuja. Reliability centred maintenance in a thermal power plant: A case study, J. International Journal of Productivity and Quality Management, 7 (2011) 209 - 228.
[4] J. W. Park, W. C. Seol. Considerations for severe accident management under extended station blackout conditions in nuclear power plants, J. Progress in Nuclear Energy, 88 (2016) 245 - 256.
[5] C. L. Zhao, A. P. Li, R. Jiang. Research on DDoS attack effect evaluation based on TOPSIS-GRA integrated evaluation method, J.Netinfo Security, 64 (2016) 40 – 46.
[6] S. Di, P. Song, W. Y. Yao, Research on reliability evaluating of wind turbine based on greyentropy AHP and TOPSIS method, J. North China Electric Power, 42 (2016) 24 - 28.
[7] S. Farhad, M. N. Hassani. J. Shahram, F. F. Hassan. Comparison of sustainability models in development of electric vehicles in Tehran using fuzzy TOPSIS method, J. Sustainable Cities and Society, 53 (2020).
[8] W. H. Sung, M. W Kim. An information entropy interpretation of photon absorption by dielectric media, J. Optics Communications, 454 (2020).
[9] R. Albert, A. Barabási. Statistical mechanics of complex networks, J. Review of Modern Physics, 74 (2002) 47 - 97.
[10] W. Y. Ju, Y. H. Li. Identification of critical lines and nodes in power grid based on maximum flow transmission contribution degree, J. Automation of Electric Power Systems, 9 (2012) 6 - 12.
[11] Y. Cai, Y. J. Cao, Y. Li. Identification of vulnerable lines in urban power grid based on voltage grade and running state, J. Proceedings of the CSEE, 34 (2014) 2124 - 2131.
[12] X. G. Chen, K. Sun. Identification of vulnerable lines in power grid based on complex network theory, J. Electric Power Automation Equipment, 26 (2006) 1 - 5.