Research on the Distribution of Balance Level Confidence Intervals for Distribution Network Engineering Samples based on Improved TOPSIS Method

Mengxiao Yi[1*], Yaqing Yang[2], Zhen Jiao[3], Jianfeng Wen[4]

{ymx6180@163.com[1*], yangyaqing5200@163.com[2], zhjiao79@126.com[3], wjf0709wjf@163.com[4]}

Electric Power Economic Research Institute, Beijing Electric Power Company, Beijing 100055, China

Abstract. With the rapid development of China's economy, the proportion of distribution network projects with voltage levels below 35KV in the whole power grid system is gradually increasing. Its construction, transformation and upgrading are also being continuously promoted. In order to meet the cost management needs of the distribution network project, this paper constructs a confidence interval measurement model of the balance level based on the improved TOPSIS method and the advance and retreat optimization strategy, and measures the project balance rate. To verify the validity and reasonableness of the model, this paper extracts the distribution network project samples of voltage levels below 35KV for case analysis.This paper can provide theoretical support for scientifically determining the control objectives of the distribution network project cost.

Keywords: Distribution network engineering; Improved TOPSIS method; Advance and retreat optimization strategy; Balance level confidence interval; Cost control

1 Introduction

Driven by the wave of new infrastructure, traditional infrastructure is facing a new round of transformation and upgrading challenges. In this context, the distribution network is no longer a mere power supply facility, but evolves into a comprehensive energy allocation platform. The deepening of the national electric power system reform has put forward higher requirements for the safety, quality, efficiency and benefits of grid operation **[1]**.

The project balance rate is a key control indicator in the process of cost management of power grid projects. It can truly reflect the accuracy of the preparation of project estimates and the degree of lean cost management. At present, the interval distribution of the balance rate of the main grid project (35kV and above) has been thoroughly studied. However, the distribution network project cost data presents discrete characteristics. The traditional simple arithmetic average method is difficult to meet the demand for lean management of distribution network project cost, because it cannot fully solve the problem of data scientificity and accuracy, thus reducing the application value of the balance rate index **[2]**.

In order to better analyze and control the cost of the distribution network project and improve the applicability of the balance rate index in the distribution network project, in this study, the

ICEMME 2023, November 17-19, Beijing, People's Republic of China

DOI 10.4108/eai.17-11-2023.2342722

TOPSIS **[3]** method, Chebyshev's Inequality **[4]** and the advance and retreat optimization strategy **[5]** are used to construct a balance level confidence interval measurement model, and the sample project data of the distribution network are randomly selected in a certain region for example analysis. This study provides theoretical support and practical analysis reference for scientifically determining the cost control objectives of distribution network projects.

2 Sample weighting method based on improved TOPSIS

The Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS) is widely used for multi-objective decision making, which is based on the concepts of positive ideal solution and negative ideal solution. Positive ideal solution refers to the ideal optimal solution, whose attribute values are the optimal values of the attributes in the evaluated solutions, while negative ideal solution is the worst solution, whose attribute values are the worst values of the attributes in the evaluated solutions. If a solution is closest to the positive ideal solution and farthest away from the negative ideal solution, it is regarded as the optimal solution, and other solutions can be ranked accordingly.

In this study, an improved TOPSIS method is proposed based on the principle of TOPSIS method for weight assignment of engineering samples. The method sets the maximum and minimum values of the standardized sample balance rate as positive and negative ideal solutions. Considering that the balance rate should be located in a reasonable interval rather than the lower the better, a distance equation is designed to ensure that samples with a balance rate close to the average advanced level receive greater weights. Since distribution network engineering has multiple voltage levels, complex network structures and diverse equipment types, its balance rate fluctuates greatly, and adopting this method can effectively reduce the influence of extreme sample data on the measurement of the balance level confidence interval, and ensure the reasonableness of the measurement results.

The specific steps of sample weight calculation based on the improved TOPSIS method in this study are shown below.

(1) Create a original balance rate data matrix A, as can be seen in **equation (1)**.

$$\mathrm{A} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{1}$$

Where, a_1 represents the balance rate of the 1st sample case; a_n represents the balance rate of the nth sample case.

(2) Standardize the original data matrix. This study uses the method of departure standardization (also known as min-max standardization), as can be seen in **equation (2)**.

$$b_{\mathrm{i}} = \frac{a_i - \min(a_{k1})}{\max(a_{k1}) - \min(a_{k1})}, k \in [1, i] \tag{2}$$

According to the above equation, the standardized data matrix B can be obtained and B is represented as **equation (3)**.

$$B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \tag{3}$$

(3) Determine the positive ideal solution a^+ and the negative ideal solution a^-, see **equations (4-5)**.

$$a^+ = max\{a_i\}\ i = 1,2,\cdots,n \tag{4}$$

$$a^- = min\{a_i\}\ i = 1,2,\cdots,n \tag{5}$$

(4) Calculate the combined distance S_i from each sample to the positive and negative ideal solutions according to the distance equation, see **equation (6)**.

$$S_i = \sqrt{0.1 \times (a_i - a^+)^2 + 1.9 \times (a_i - a^-)^2} \tag{6}$$

(5) Finally, the weight of each sample ω_i is calculated based on the combined distance of each sample, see **equation (7)**.

$$\omega_i = \frac{\frac{1}{S_i}}{\sum_1^n \frac{1}{S_i}} \tag{7}$$

The engineering samples can be weighted based on the calculated weights obtained for each sample, providing the basis for subsequent balance level confidence interval measurement models.

3 Modeling of balance level confidence intervals

Chebyshev's Inequality is a fundamental inequality in probability theory. The inequality gives a lower bound on the probability of any random variable within a certain distance from its mean. Specifically, for any random variable with finite variance, Chebyshev's Inequality can be expressed in **equation (8)**.

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \tag{8}$$

Where P denotes the probability, X is the random variable, μ is the mean of the random variable, σ is the standard deviation of the random variable, and k is any positive real number greater than 1. In this paper, the distribution network engineering balance level confidence interval measurement model is constructed according to Chebyshev's Inequality. According to Chebyshev's Inequality to construct the distribution network engineering balance level confidence interval expression is shown in **equation (9)**.

$$\varphi = \{x|\mu - k\sigma \leq x \leq \mu - k\sigma\} \tag{9}$$

Where μ is the sample weighted mean; k is the interval coefficient; and σ is the standard deviation.

From Chebyshev's Inequality, it can be obtained that the arbitrary distribution data contains at least C of the whole sample data within k standard deviations.

Chebyshev's Inequality in the case of unknown distribution of data, usually the data within the mean k times the standard deviation percentage example the following relationship exists,see **equation (10)**.

$$C = 1 - \frac{1}{k^2} \tag{10}$$

In order to ensure the scientificity of the determination of the confidence interval of the balance rate, the proportion of the sample estimation interval is set to $C = 80\%$ is set. According to **equation (10)**, it can be obtained that when at least 80% of the sample points are in the interval, $k = 2.2361$. Therefore, this value is used as the initial value of the algorithm solution in this paper in order to speed up the algorithm's optimization search and solution process.

In this paper, the problem is solved using the advance and retreat optimization strategy. It‘s a numerical algorithm that determines the search interval and guarantees an approximate single-peak property, which can be used to solve multi-objective decision-making in complex problems. The logic of the operation is shown in **Table 1**.

Table 1. Advance and retreat optimization strategy operation logic

Assuming that	If	Then
$F(x)$ is a single-valley function and the extreme point is located at $[a,b]$	$\forall x_1, x_2 \in [a,b], f(x_1) < f(x_2)$	The search interval for the minima is $[a, x_2]$
	$\forall x_1, x_2 \in [a,b], f(x_1) > f(x_2)$	The search interval for the minima is $[x_1, b]$

Therefore, given an initial point x_0 and an initial search step h, $f(x_0 + h)$ is calculated by first searching one step forward with the initial step. The operation mechanism is shown in **Fig 1**.

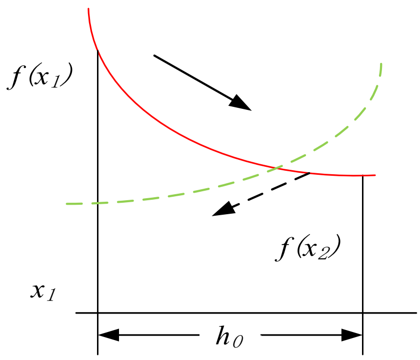
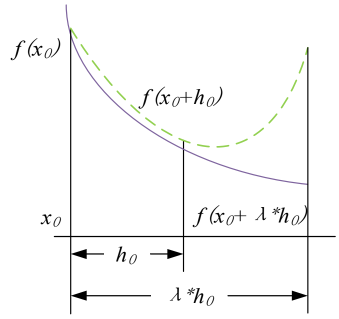
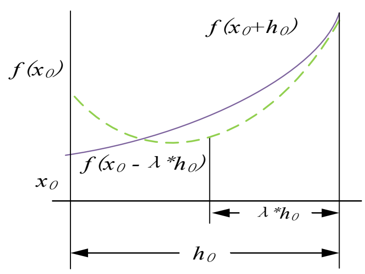


Fig 1. Algorithm of Advance and retreat optimization strategy

The calculation of $f(x_0 + h)$ produces two results, $f(x_0) > f(x_0 + h)$ and $f(x_0) < f(x_0 + h)$. For these two cases, different calculations need to be performed, and the calculation is iterated until the coefficients satisfying the corresponding conditions are determined, thus determining the new search interval. The specific calculation method and process diagram are shown in **Table 2**.

Table 2. Operation rules of advance and retreat optimization strategy

If	Then	Illustration of the calculation process
$f(x_0) > f(x_0 + h)$	The search interval is $[x_0, y]$, where y is the parameter to be searched, and forward calculating $f(x_0 + \lambda h)$, with λ as an incremental factor and $\lambda > 1$, until the iteration yields the unique λ^* such that $f(x_0 + \lambda h) < f(x_0 + \lambda^* h))$, which determines the new search interval as $[x_0, x_0 + \lambda^* h]$.	
$f(x_0) < f(x_0 + h)$	The search interval is $[y, x_0 + h]$, where y is the parameter to be searched, and backward calculating $f(x_0 - \lambda h)$, with λ as an decrement factor and $0 < \lambda < 1$, until the iteration yields the unique λ^* such that $f(x_0 - \lambda^* h) < f(x_0))$, which determines the new search interval as $[x_0 - \lambda^* h, x_0 + h]$.	

The optimal interval coefficient k is obtained iteratively using the advance and retreat optimization strategy, and the proportion of the total number of samples falling into the interval is used as the basis for discrimination, and the objective function is constructed using the advance and retreat optimization strategy to optimize the interval coefficient k iteratively, in order to find the unconditional extreme value $minF(\lambda)$ of the objective function, and the objective function is expressed as **equation (11)**.

$$F(k) = \left| \frac{num(\mu - k\sigma \leq x \leq \mu - k\sigma)}{num(z)} - 80\% \right| \tag{11}$$

The optimization process of the advance and retreat optimization strategy is shown in Figure 1. In accordance with the 80% interval estimation requirements to get $k_0 = 2.24$, iterative search to calculate the interval coefficient corresponding to the project balance level confidence interval, to determine whether the sum of the weights of the samples falling into the interval to meet the 80% of the sum of the weights of the total samples. If it is satisfied, the interval coefficient is output and the iteration is completed; if it is not satisfied, the step size is modified and the interval calculation process is repeated. The iterative process of the advance and retreat optimization algorithm is shown in **Fig 2**.

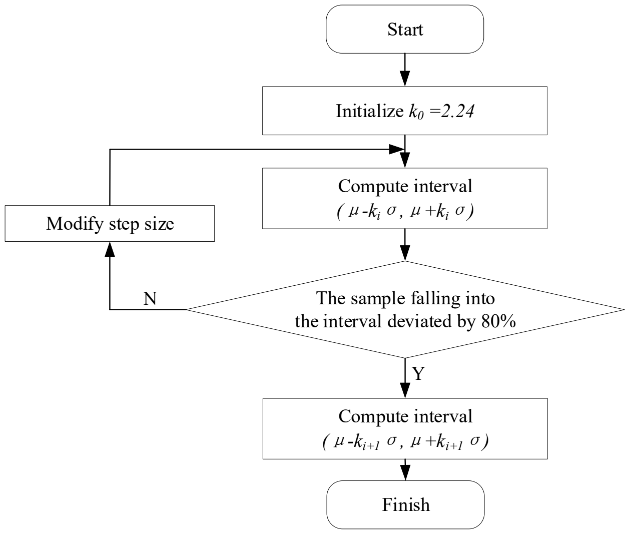


Fig 2. Iterative process of the advance and retreat optimization algorithm

4 Example analysis

4.1 Data base

This example analyzes the cost data of each investment stage of the partial distribution network project completed and put into operation by a power company in a certain region of China in 2021, and studies the balance level confidence interval distribution. The sample data contains a total of 170 projects, as shown in **Fig 3**.

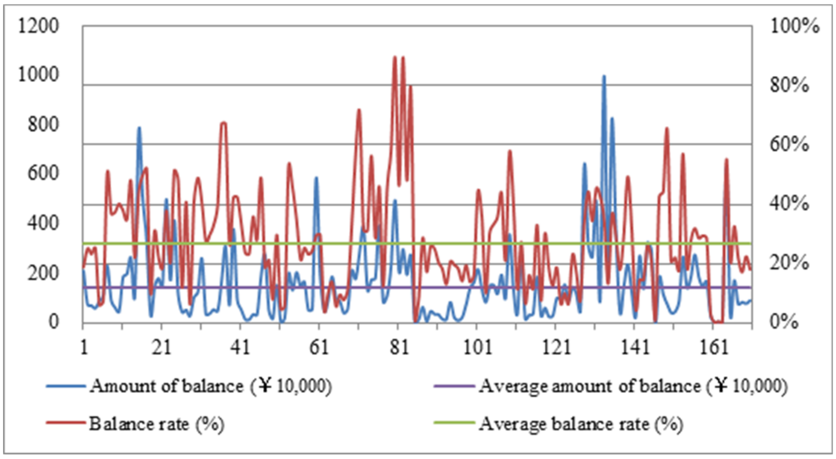


Fig 3. Statistical chart of dynamic investment balances and ratios for sample projects

As can be seen from the above figure, the highest balance amount in the sample data is 9,992,000 yuan, and the project with the smallest balance amount has no balance; the highest balance rate is 89.66%. The average of balance amount and balance rate of the distribution network project is 1,425,600 yuan and 26.82% respectively. Due to the high degree of sample dispersion, the average level is difficult to directly reflect the reasonable range of the balance level of the distribution network project, and cannot meet the needs of cost management accuracy.

4.2 Calculation of balance level confidence intervals for distribution network project

(1) Sample weighting calculation

First, based on the sample weighting method of improved TOPSIS, the sample data are weighted to reduce the impact of discrete data on the measurement results, while enabling the measurement results to reflect the average advanced level, the specific process is as follows.

1) Calculate the balance rate of each sample case.

2) Establish the original data matrix A. Establish a single-column matrix of 184 rows for the base data of the 184 samples used in this paper, representing the balance rate of each project respectively, and obtain the original data matrix A expressed as **equation (12)**.

$$\mathrm{A} = \begin{bmatrix} 18.71\% \\ 24.94\% \\ \vdots \\ 17.98\% \end{bmatrix} \tag{12}$$

3) Standardize the original data matrix A. The data matrix B obtained after standardized processing is represented as **equation (13)**.

$$\mathrm{B} = \begin{bmatrix} 0.2002 \\ 0.2669 \\ \vdots \\ 0.1924 \end{bmatrix} \tag{13}$$

4) As the data were standardized for deviation, the positive and negative ideal solutions a^+ and a^- were determined to be 1 and 0 respectively.

5) The combined distance of each sample to the positive and negative ideal solutions S_i was calculated according to the distance equation. The calculated distance of each sample to the ideal solution is shown as **equation (14)**.

$$S = (0.4245,0.5102,\cdots,0.4723,0.4154) \tag{14}$$

6) Finally calculate each sample weight ω_i based on the combined distance of each sample. The calculated weight of each sample see **equation (15)**.

$$W = (0.00682,0.00567,\ldots,0.00613,0.00697) \tag{15}$$

(2) Balance level confidence interval measurement

According to the method of determining the project balance interval based on the advance and retreat optimization algorithm, the initial value interval ϕ of the cost balance of the grid primary substation project is obtained from Chebyshev's Inequality as **equation (16)**.

$$\phi = \{x | 0.1803 - 0.1266k \le x \le 0.1803 + 0.1266k\} \tag{16}$$

Then the objective function of the advance and retreat optimization process see **equation (17)**.

$$F(k) = \left| \frac{\text{num}(|0.1803 - 0.1266k \leq x \leq 0.1803 + 0.1266k)}{170} - 80\% \right| \quad (17)$$

Find the interval coefficient k so that $F(k)$ is minimized. Set the initial step size $h_0 = -0.1, \varepsilon = 0.005, k_0 = 2.2361$. The optimization process of the advance and retreat optimization method is shown in **Table 3**. It can be seen from the table that, when iterating for 11 times, $|h| < \varepsilon$, iteration terminates, $k_x = 0.6361$, and thus the confidence interval see **equation (18)**.

$$\phi_1 = \{x|9.98\% \leq x \leq 26.08\%\} \quad (18)$$

Table 3. Optimization of interval parameters of the grid primary substation project

Iteration number	Moving direction	h_i	k_i	Search result
0	-	-	2.2361	-
1	Forward	-0.1	2.1361	Succeed
2	Forward	-0.2	1.9361	Succeed
3	Forward	-0.4	1.5361	Succeed
4	Forward	-0.6	0.9361	Succeed
5	Forward	-0.8	0.1361	Fail
6	Backward	0.1	0.2361	Succeed
7	Backward	0.2	0.4361	Succeed
8	Backward	0.4	0.8361	Fail
9	Forward	-0.1	0.7361	Succeed
10	Forward	-0.2	0.5361	Fail
11	Backward	0.1	0.6361	Succeed

In summary, the balance level confidence interval of distribution network projects obtained based on the method of this paper is [9.98%, 26.08%]. Among the sample projects, a total of 74 projects have a balance rate higher than the maximum value of the confidence interval, accounting for 52.86% of the total sample. Therefore, there is still much room for improvement in the lean cost control of distribution network projects.

5 Conclusion

Based on the cost data of each investment stage of the distribution network project in 2021, this paper constructs the research method of the balance level confidence interval of the distribution network project by using the improved TOPSIS method and the advance and retreat optimization strategy, and finally derives the balance level confidence interval of the engineering. In the process of engineering investment review and benefit evaluation, the balance level confidence interval of engineering is used as an indicator reference, which can reduce the dependence on experts' experience and improve the applicability and accuracy of the balance rate indicator. When the accumulated historical engineering sample data reaches a certain level, the method can calculate the more representative confidence interval value of the engineering balance level, which provides a reliable analytical basis for the investment control and benefit evaluation of regional distribution network projects and has practical significance.

References

[1] L. Yuanhong et al., Big Data Based Analysis Between Power Distribution Network Reliability Parameters and Economic and Social External Environment, 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), , pp. 528-531, (2021)
[2] A. Vafamehr and M. E. Khodayar, Operation of Distribution Networks with Volatile Supply and Controllable Data Center Demand, IEEE/PES Transmission and Distribution Conference and Exposition (T&D), pp. 1-5, (2018)
[3] A. Elhassouny and F. Smarandache, Neutrosophic-simplified-TOPSIS Multi-Criteria Decision-Making using combined Simplified-TOPSIS method and Neutrosophics, 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), pp. 2468-2474,(2016)
[4] J. Ding, Z. Chen, X. He and Y. Zhan. Clustering by finding density peaks based on Chebyshev's inequality. 2016 35th Chinese Control Conference (CCC), pp. 7169-7172, (2016)
[5] Jun qi Zhang, Zhong min Xiao, Ying Tan and Xin gui He. Hybrid particle swarm optimizer with advance and retreat strategy and clonal mechanism for global numerical optimization. 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), pp. 2059-2066, (2008)