Management and development performance assessment for electric distribution company based on data mining

Shen Zeyuan¹, Wang Zheng¹, Zhao Haibo¹, Song Xiaojun¹

¹Economic and Electrical Research Institute of Shanxi Electrical Power Company of SGCC, Planning Review Center, Taiyuan, People’s Republic of China

E-mail: szy33519@163.com

Abstract: In this study, the Statistical Product and Service Solutions software is applied to analyse the massive data of electric distribution companies. A comprehensive evaluation of grid development and production and operation of basic electric distribution companies is the key to a company's investment and development strategies. This study proposes a comprehensive evaluation index system for electric distribution companies. In the method, the weight of each index is calculated using the improved analytic hierarchy process based on the Delphi method. Then, according to the actual operation situation of each enterprise, the differential weight of various indices is constructed, and the comprehensive evaluation and score of differentiation for electric distribution companies are realised, which can be used for locating the weak line of the power grid of each enterprise and putting forward an investment strategy of the power grid. Through the demonstration application of 98 electric distribution companies in Shanxi Province of China, this method exhibits a promotion of value and accuracy in carrying out a comprehensive evaluation for electric distribution companies.

1 Introduction

Electric distribution companies are directly responsible for the operation and management of distribution networks. For a long time, influenced by various factors such as management system, personnel structure, equipment level, operation principle, and regional differences, electric distribution companies have been lagging in distribution network construction and management, planning, and operations. Therefore, a comprehensive evaluation based on electric distribution companies is urgently needed. Through the evaluation method, all aspects of production, operation, and power grid development can be evaluated and the weakness can be located, which provide a foundation for distribution companies to formulate investment strategies and improve investment efficiency.

To achieve a comprehensive evaluation of the power grid, most traditional evaluation methods [1, 2] are based on the actual operation and development indicators of the grid, using the entropy weight method and dynamic comprehensive evaluation method to determine the weight of each index and carry out index scoring. A previous study [3] used the analytic hierarchy process (AHP) to carry out a comprehensive evaluation of the rural low-voltage power grid. In addition, a previous study [4] has proposed a method and a procedure for grid planning evaluation by constructing a multi-dimensional grid evaluation index system. The literatures [5, 6] have conducted a comprehensive evaluation of the power grid through data mining methods such as principal component analysis (PCA) and system clustering. Most studies have described the methods and processes for comprehensive evaluation of power grids, but lack the multi-level evaluation for the development of the power grid, production, and operation, especially the diversity of the power grid.

This paper uses Statistical Product and Service Solutions (SPSS) [7, 8], through PCA, K-means clustering, correlation analysis, and systematic clustering, to reduce the dimension of the massive data and indicators of electric distribution companies and to achieve enterprise classification. Then, Specific, Measurable, Attainable, Relevant, Trackable (SMART) guidelines are used to construct the evaluation indicator system for basic electric distribution companies, and all indicators are normalised using different standardised strategies. On this basis, an improved AHP method based on the Delphi method is used to calculate the index weights, and based on the actual situation of each company, different weights are proposed. Finally, a comprehensive evaluation of basic electric distribution companies is achieved. The differentiated weights can better reflect the actual situation and characteristics of the power grid in each company, which conforms to the nature and rigorousness of evaluation and improves the application value and theoretical value.

2 Data mining

Data mining, also known as knowledge discovery in the database, is a complex process that extracts knowledge of unknown, valuable patterns or laws from large amounts of data. Data mining mainly includes seven steps: data cleansing, data integration, data selection, data transformation, data mining, pattern assessment, and knowledge representation. Common data mining methods include mathematical statistics, neural networks, genetic algorithms, and decision tree induction [9–12].

This paper is based on the massive data of 98 basic electric distribution companies in Shanxi Province of China, using SPSS software as a data analysis platform, PCA, K-means clustering, and correlation analysis. In-depth data mining and data analysis were performed on 198 grid development, production, and operation indicators, and the number ultimately reduced to 40 indicators. The specific mining process is shown in Fig. 1.

Through deep data mining and analysis, comprehensive utilisation of K-means clustering and PCA, the 98 electric distribution companies in Shanxi Province is divided into five categories. The clustering result fully takes into account the economic development situation of each region and the development of power grids. From the classification results, it is possible to trace back the main component factors of each enterprise's decision categories, thereby positioning the weaknesses of each company's power grid development.

3 Evaluation system

3.1 Evaluation index system construction

The evaluation index system constructed in this study uses SMART criteria to ensure that the indicator system covers all aspects and processes of the grid development and production operations of...
basic electric distribution companies. It can fully reflect the socioeconomic, equipment operation, power grid production, grid-connected power generation, operation and management, and energy saving and emission reduction properties of basic power supply grids. On the basis of in-depth analysis of practical issues, various factors that affect the development of the power grid and production and operation are broken down into several levels. The factors on the same level are subordinated to those on the upper level or have an impact on the superior factors. At the same time, it also controls the factors of the next level or is influenced by the factors of the lower level [13, 14]. At the end, it is determined that grid development, production, and operation are the primary targets, and a hierarchical structure model of the electric distribution companies is established, as shown in Fig. 2.

3.2 Data standardisation strategy
As there are many types of indicator data for basic electric distribution companies, the goals and requirements of different indicators in the power grid are inconsistent. Therefore, different standardised calculation methods need to be adopted according to the types of indicators. For the indicators involved in the evaluation index system of electric distribution companies, this paper mainly adopts three standardised processing strategies: trapezoidal processing strategy, semi-triangular processing strategy, and segmented processing strategy [15].

(i) Trapezoidal processing strategy: The trapezoidal processing strategy mainly involves 24 indicators such as the load ratio, the average power supply line radius, the N–1 pass rate, the insulation
rate, and the number of power grid accidents. Taking the load ratio as an example, in the ‘Technical Guidelines for Planning and Designing Distribution Network’, the 110–35 kV load ratio must be controlled between 1.8 and 2.2. First, the load ratio is converted into the normalised value of the percentage system; then, according to the range of the load ratio interval, the trapezoidal processing function of the load ratio is established, as shown in Fig. 3.

The specific mapping of the trapezoidal processing function for the load/load ratio is shown in Table 1.

Table 1  Calculation standards for capacity/ratio specifications

| Function type                  | Actual value range | Normalised value |
|-------------------------------|--------------------|------------------|
| trapezoidal processing functions | $x \in [0, 1)$     | $f(x) = 0$       |
|                               | $x \in [1, 1.8)$   | $f(x) = 125 \times (x - 1)$ |
|                               | $x \in [1.8, 2.2)$ | $f(x) = 100$    |
|                               | $x \in [2.2, 3)$   | $f(x) = 125 \times (3 - x)$ |
|                               | $x \in [3, + \infty)$ | $f(x) = 0$       |

where $x$ is the actual value and $f(x)$ is the normalised value.

![Fig. 3 Trapezoidal processing function](image)

(iii) Semi-triangular processing strategy: The semi-triangular processing strategy mainly involves 12 indicators such as the coverage of electricity information acquisition system, automatic distribution coverage, and smart meter coverage. Taking the coverage rate of the electricity information acquisition system as an example, the actual value of the electricity information collection system coverage is first converted into the standardised value of the percentage system. The coverage rate of the electricity information acquisition system is a positive indicator. We then establish a monotonously increasing half-triangular processing function for the coverage of the power usage information acquisition system, as shown in Fig. 4.

The specific mapping of the trapezoidal processing function for the load/load ratio is shown in Table 1.

(ii) Segmentation strategy: The segmentation strategy mainly involves the following four indicators: asset-liability ratio, personal injury, and death. Taking personal casualties as an example, the actual value of personal accident casualties is first converted into a standardised value of the percentage system. Personal accident casualties are negative-type indicators and a segmented processing function is established, as shown in Fig. 5.

The specific mapping of the trapezoidal handler function for personal accident casualties is shown in Table 3.

![Fig. 4 Triangulation function for mining system coverage. M equals to 1](image)

Table 2  Standard calculation standards for coverage of use systems

| Function type                  | Actual value range | Normalised value |
|-------------------------------|--------------------|------------------|
| single increase in triangle processing function | $x \in [0, 1]$  | $f(x) = 100x$ |

Among them, $x$ is the actual value, and $f(x)$ is the normalised value.

![Fig. 5](image)

4 Evaluation system

4.1 Benchmark weight calculation

In this study, the Delphi method is used to improve the AHP [16, 17], and several expert experience opinions are synthesised to calculate the index weight parameters in the index system. The main process is as follows.

(i) Each of the $n$ experts is asked to score the indicators at each level to form $n$ judgment matrices

\[
R_i = \begin{bmatrix}
    x_{i1} & x_{i2} & \cdots & x_{im} \\
    x_{i1} & x_{i2} & \cdots & x_{im} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{i1} & x_{i2} & \cdots & x_{im}
\end{bmatrix}
\]

(ii) Singular values are removed (big difference matrix). Assuming that the $k$th expert establishes the judgment matrix as
Table 3 Calculation standards for personal accident casualties

| Function type | Actual value range | Normalised value |
|---------------|--------------------|-----------------|
| single increase in triangle processing function | $x \in [0, 1]$ | $f(x) = 100x$ |

Among them, $x$ is the actual value and $f(x)$ is the normalised value.

4.2 Differentiated weight calculation

There are many basic electric distribution companies in Shanxi Province. Due to the different levels of economic development, environment, and climate in various regions, different enterprises have different levels of power grid development. To improve the accuracy of the evaluation method, this paper modifies the benchmark weights, fully combines the characteristics of the development of each region, and presents the calculation results of differential weights for various regions. Taking the correction of the weights of secondary indicators of power grid development as an example, the specific calculation process is described as follows:

(i) Build a differentiated scoring matrix: According to the recommendations of the expert group, different categories of the same indicator are scored, with a scoring standard of 1–5. When the indicator has a large weight in a certain category, 5 points are assigned, as shown in Table 5.

(ii) Calculate the index contribution matrix: According to the differential scoring matrix, the scores of the same category of regions are normalised and the contribution of each index to the region is calculated and defined as $\theta_{ij}$, indicating the degree of contribution of the $i$th index to the $j$th region, as shown in Table 6.

(iii) Calculate the weight correction matrix: According to the index contribution matrix, the difference correction matrix of each index in different categories is calculated by combining the reference weights $\phi_i$ of each index. The formula is $\omega_{ij} = \phi_i \times (0.5 + \theta_{ij})$. The details are shown in Table 7.

(iv) Calculate the differential weight: According to the weight correction matrix, the normalisation process can be applied to obtain differential weights applicable to different types of regions, as shown in Table 8.

5 Applications

Using a comprehensive evaluation method for basic electric distribution companies based on differentiated weights, 98 basic electric distribution companies in Shanxi were evaluated and scored. According to the scores of various indicators, we can intuitively locate the weaknesses of each company’s power grid development.

Taking Yangqu county in the first category and Pingshun county in the fourth category as an example, through retrospective analysis of the clustering results, it was found that the fourth type of power supply enterprises is mostly poor areas, the development of power grids is slow, and the power supply capacity still needs to be significantly increased. Therefore, the power supply capacity has the highest weight, i.e. 0.295, which is higher than the reference weight of 0.251 and other categories of differential weights. While the first type of enterprises is mostly in areas where the power grid development...
develops rapidly, the power supply capacity has basically met the power supply demand, but the grid structure still needs to be strengthened. Therefore, the grid structure has the highest weight, i.e. 0.278, which is higher than the benchmark weight of 0.276 and other types of differential weights.

Table 4: Weight of power grid development indicator evaluation system for Shanxi primary power supply enterprises

| First-level indicators | Weights | Secondary indicators | Weights | Third-level indicators | Weights |
|------------------------|---------|----------------------|---------|------------------------|---------|
| development of the power grid | 0.5     | capability of power supply | 0.251   | capacity ratio          | 0.313   |
|                        |         | average distribution capacity per household |         |                         |         |
|                        |         | overload ratio of the line |         |                         |         |
|                        |         | light-load ratio of the line |         |                         |         |
|                        |         | main transformer (power distribution) overload ratio |         |                         |         |
|                        |         | main transformer (power distribution) light-load ratio |         |                         |         |
| the structure of the power grid | 0.276   | N–1 Line pass rate | 0.348   |                         |         |
|                        |         | N–1 Main transformer pass rate | 0.424   |                         |         |
| technical equipment | 0.118   | the average power supply radius of the line | 0.228   |                         |         |
|                        |         | average power supply reliability | 0.357   |                         |         |
|                        |         | fault outage time | 0.325   |                         |         |
|                        |         | planned outage time | 0.175   |                         |         |
|                        |         | comprehensive voltage qualification rate | 0.143   |                         |         |
| power quality | 0.243   | electricity information acquisition system coverage | 0.363   |                         |         |
|                        |         | distribution automation coverage | 0.322   |                         |         |
|                        |         | smart meter coverage | 0.315   |                         |         |
| intelligent | 0.112   |                         |         |                         |         |

Table 5: Rating of power grid development indicators

| Index                        | Category 1 | Category 2 | Category 3 | Category 4 | Category 5 |
|------------------------------|------------|------------|------------|------------|------------|
| power supply capability      | 2          | 3          | 1          | 5          | 4          |
| structure of the power grid  | 3          | 3          | 2          | 3          | 2          |
| technical equipment          | 5          | 2          | 1          | 3          | 4          |
| power quality                | 3          | 4          | 4          | 2          | 2          |
| intelligent                  | 5          | 4          | 3          | 2          | 2          |

Table 6: Index contribution matrix

| Index                        | Category 1 | Category 2 | Category 3 | Category 4 | Category 5 |
|------------------------------|------------|------------|------------|------------|------------|
| power supply capability      | 0.1111     | 0.1875     | 0.0909     | 0.3333     | 0.2857     |
| the structure of the power grid | 0.1667   | 0.1875     | 0.1818     | 0.2000     | 0.1429     |
| technical equipment          | 0.2778     | 0.1250     | 0.0909     | 0.2000     | 0.2857     |
| power quality                | 0.1667     | 0.2500     | 0.3636     | 0.1333     | 0.1429     |
| intelligent                  | 0.2778     | 0.2500     | 0.2727     | 0.1333     | 0.1429     |

Table 7: Indicator correction matrix

| Index                        | Category 1 | Category 2 | Category 3 | Category 4 | Category 5 |
|------------------------------|------------|------------|------------|------------|------------|
| power supply capability      | 0.1536     | 0.1728     | 0.1485     | 0.2095     | 0.1975     |
| the structure of the power grid | 0.1841   | 0.1899     | 0.1838     | 0.1934     | 0.1776     |
| technical equipment          | 0.0919     | 0.0739     | 0.0698     | 0.0827     | 0.0929     |
| power quality                | 0.1618     | 0.1820     | 0.2096     | 0.1537     | 0.1560     |
| intelligent                  | 0.0868     | 0.0837     | 0.0862     | 0.0707     | 0.0717     |

Table 8: Differentiated weight design

| Index                        | Category 1 | Category 2 | Category 3 | Category 4 | Category 5 |
|------------------------------|------------|------------|------------|------------|------------|
| power supply capability      | 0.251      | 0.225      | 0.246      | 0.211      | 0.295      |
| structure of the power grid  | 0.276      | 0.278      | 0.270      | 0.268      | 0.272      |
| technical equipment          | 0.118      | 0.132      | 0.105      | 0.099      | 0.117      |
| power quality                | 0.243      | 0.237      | 0.259      | 0.298      | 0.216      |
| intelligent                  | 0.112      | 0.128      | 0.119      | 0.123      | 0.100      |

The comprehensive scores of Yangqu County and Pingshun County are shown in Table 9. Fig. 6 is a radar chart of the scores of Yangqu County and Pingshun County, which can clearly locate the weaknesses of the company, and then go back to the specific three-level indicators, find out the root cause, and guide the investment focus and investment direction of the distribution companies in the
future. Therefore, the scores of Yangqu County and Pingshun County indicate the accuracy and rationality of the method of differential weights, which can reflect the differences between electric distribution companies in various regions to a certain extent, making the evaluation results more accurate.

**6 Conclusion**

A comprehensive evaluation of grid development and production and operation of basic electric distribution companies is the key to a company's investment and development strategies. Through a comprehensive evaluation of the electric distribution companies proposed in this paper, it can clearly locate the problems existing in the operation of various enterprises and the development of power grids, identify the direction for the company's next-step investment focus, improve the availability of power grid investments, ensure the efficiency of investment, and further implement the accurate investment.

**Table 9** Scoring table for differentiated evaluation of distribution companies in Yangqu county and Pingshun county

| Enterprise          | Category | Power supply capacity | Grid structure | Technical equipment | Power quality | Intelligent operating results | Development investment | Safe production | Power service | Score  |
|---------------------|----------|-----------------------|----------------|--------------------|--------------|-------------------------------|------------------------|-----------------|--------------|--------|
| Yangqu county       | I        | 50.69                 | 99.57          | 98.81              | 100          | 100                           | 70.29                  | 60              | 100          | 90.05  |
| Pingshun county     | IV       | 76.01                 | 83.28          | 96.67              | 28.57        | 100                           | 90.26                  | 81.84           | 100          | 84.62  |

Fig. 6 Radar of power grid development and production operation of power grid enterprises in Yangqu county and Pingshun county

**7 References**

[1] Hairui, Z., Dong, H., Yu, L., et al.: ‘Smart grid evaluation based on anti-entropy weight method’, *Power Syst. Prot. Control*, 2012, 40, (11), pp. 24–29

[2] Dong, H., Zheng, Y., Yiqun, S., et al.: ‘Dynamic assessment method for smart grid based on system dynamics’, *Autom. Electr. Power Syst.*, 2012, 36, (3), pp. 16–21

[3] Yang, C., Hanhui, M., Li, Z., et al.: ‘A comprehensive evaluation of new rural low-voltage distribution networks based on analytic hierarchy process’, *Power Syst. Technol.*., 2007, 31, (8), pp. 68–72

[4] Jun, X., Yanyan, C., Jinnmin, W., et al.: ‘A hierarchical performance assessment method on the distribution network planning’, *Autom. Electr. Power Syst.*, 2008, 32, (15), pp. 36–40

[5] Weixing, L., Peng, W., Zhimin, L., et al.: ‘Reliability evaluation of complex radial distribution systems considering restoration sequence and network constraints’, *IEEE Trans. Power Deliv.*, 2004, 19, (2), pp. 753–758

[6] Xinhua, G., Zheng, Y.: ‘Comprehensive assessment of smart grid construction based on principal component analysis and cluster analysis’, *Power Syst. Technol.*, 2013, 37, (8), pp. 2238–2243

[7] Runlong, H.: ‘Statistical analysis of data-SPSS principle and application’ (Higher Education Press, Beijing, 2010)

[8] Shanshan, Z.: ‘CPI analysis based on system clustering in SPSS’, Xinxiang, Henan Normal University, 2013

[9] Qingshan, X., Wendi, W., Zhangsui, L., et al.: ‘Establishment and application of EMI indicator system orienting to massive industrial data mining’, *Electr. Power Autom. Equip.*, 2015, 35, (7), pp. 15–21

[10] Yaqi, S., Guoliang, Z., Yongli, Z.: ‘Present status and challenges of big data processing in smart grid’, *Power Syst. Technol.*, 2013, 37, (4), pp. 927–935

[11] Dongxia, Z., Xin, M., Liping, L., et al.: ‘Research on development strategy for smart grid big data’, *Proc. CSEE*, 2015, 35, (1), pp. 2–12

[12] Peijian, W.: ‘Dynamic data center operations with demand-responsive electricity prices in smart grid’, *IEEE Trans. Smart Grid*, 2012, 3, (4), pp. 1743–1754

[13] Zhentao, H., Zhiwei, H., Shaoyun, G., et al.: ‘A comprehensive evaluation system of urban distribution network’, *Power Syst. Technol.*, 2012, 36, (8), pp. 95–99

[14] Xinhua, G., Zheng, Y.: ‘The smart grid's evaluation index system with technology maturity characteristic’, *South. Power Syst. Technol.*, 2014, 5, pp. 8–12

[15] Guan, L., Qizong, W.: ‘Research on data standardization in comprehensive evaluation based on consistent result’, *Math. Pract. Theory*, 2014, 3, pp. 8–12

[16] Gantz, J., Reinsel, D.: ‘Extracting value from chaos’. Proc. IDC iView. Framingham, USA: [n.s.], 2011, pp. 1–12

[17] Zhenghang, H., Zhijing, Y., Shaohua, L., et al.: ‘The contribution of double-fed wind farms to transient voltage and damping of power grids’, *Power Syst. Prot. Control*, 2015, 22, (1), pp. 43–44