Correlation Analysis of Production and Operation Indicators of Power Grid Enterprises Based on XGBoost Model

Liyu Xia[1,*], Yuheng Sha[2], Zhuo Song[2], Hong Liu[3], Wan He[1]

[*]Corresponding author's e-mail: xialiyu@sgeri.sgcc.com.cn

State Grid Energy Research Institute Co., Ltd., Beijing, 102209, China[1]
State Grid Corporation of China Co., Ltd., Beijing, 100031, China[2]
State Grid Hebei Electric Power Co., Ltd. Information and Communication Branch, Shijiazhuang, 050013, China[3]

Abstract. It is of great significance to establish a set of key indicators for production and operation of power grid enterprises, clarify the relationship between production indicators and operation indicators, and provide technical methods for analyzing the overall changes in production and operation caused by abnormal changes in a certain indicator of power grid enterprises. This paper combs the key indicators of production and operation of power grid enterprises and forms a set of key indicators of production and operation. Based on the variable importance judgment method of XGBoost model, judge the strength of the correlation relationship between indicators, and form a network diagram of the correlation relationship between indicators. The research shows that there is a linkage relationship between production indicators and operation indicators of power grid enterprises. Basic indicators mainly affect the power supply or sales of power grid enterprises, and less affect production or operation indicators. Power grid enterprises can better predict and manage production and operation activities according to the relationship between key indicators.

Keywords. Correlation analysis, Production and operation, Indicator management

1. Introduction

The power grid enterprises takes the investment, construction and operation of the power grid as their core business, and undertake the basic mission of ensuring safe, economic, clean and sustainable power supply. They are large backbone enterprises related to the lifeline of the national economy and national energy security. Production and operation is an important work for power grid enterprises to continuously provide power supply services and ensure the sustainable operation status. High-quality production and operation management is of great significance to ensure the implementation of enterprise strategy and planning, and to achieve optimal overall efficiency and efficiency. At present, the internal and external environment faced by power grid enterprises is undergoing profound and complex changes. Power grid enterprises must attach great importance to production and operation management, and further optimize indicators and design models. They must pay attention to mining the linkage between indicators, and pay attention to the core role of key indicator management in production and operation.

ICBBEM 2023, May 19-21, Hangzhou, People's Republic of China

DOI 10.4108/eai.19-5-2023.2334366

The academia and industry have carried out a lot of research or practice on the optimization of enterprise production and operation. Quanjie et al. (1998)[1] established the membership function of risk indicators, and carried out risk measurement through fuzzy comprehensive evaluation to analyze and evaluate the production and operation risks of enterprises. Wei Jian (2015)[2] constructed a new comprehensive plan management indicator system and evaluated the new indicator system. Zhao Ping (2015)[3] started with the interpretation of cost indicators related to decision-making, briefly outlined the four decision-making analysis methods commonly used in short-term business decisions of enterprises. Zhang Shenggen (2017)[4] believed that only by taking cost reduction and efficiency increase as a long-term mechanism to improve management level and profitability can we effectively improve the comprehensive competitiveness and profitability of enterprises. Ning Yuling (2017)[5] believed that through the production and operation indicators of thermal power plants, the current production and operation status of the thermal power industry can be grasped, and data and information support can be provided when the state regulates relevant policies. Gong Wenqing (2019)[6] believed that the establishment of performance evaluation indicators should be based on financial indicators, and can also be based on non-financial related data indicators. Zheng Chen et al. (2021)[7] refined key management indicators on the basis of summarizing the reform practice and specific practices of multi-dimensional lean management of State Grid. Shen Yuewei (2022)[8] analyzed the impact of substation integrated automation transformation, line installation of intelligent online monitoring system and other means on the production and operation of power enterprises. Han Yiming et al. (2022)[9] established a comprehensive evaluation index system suitable for the development and operation of power grid, and calculated the weight of each index based on the improved CRITIC-entropy weight method.

This paper will sort out the key indicators of production and operation of power grid enterprises and form a set of key indicators of production and operation. Based on the variable importance judgment method of XGBoost model, judge the strength of the correlation relationship between indicators, and form a network diagram of the correlation relationship between indicators. On the one hand, judge the relative importance of various indicators, and on the other hand, describe the relationship between indicators.

2. Indicators and methods

2.1. Key indicators selection

Various indicators come from the accumulation of data in the current power grid enterprise information management system. Select key indicators according to the two dimensions of production and operation, and establish an indicator set. At the same time, considering the impact of the economic and social environment of the region where the enterprise is located on the production and operation, the external macro environmental indicators are introduced.

In terms of production indicators, power supply volume, line loss rate, capacity expansion are selected. Power supply refers to the total amount of electricity delivered to users through the substation, including the power loss in transmission. The line loss rate is the percentage of the line loss energy in the power supply. The line loss energy includes all electric energy losses from the primary side of the main transformer of the power plant (excluding auxiliary power) to the user's electricity meter. The capacity expansion is the sum of the rated capacity or rated power of all electrical equipment related to the development of new power users.

In terms of operation indicators, power generation volume, total assets, operating income, total profit are selected. Power generation volume refers to the amount of electric energy generated by the generator through energy conversion, including thermal power generation, hydropower generation, nuclear power generation and other power generation of all electric power industries, captive power plants, rural small power plants. Total assets are all assets owned or controlled by an enterprise, including current assets, long-term investments, fixed assets, intangible and deferred assets, and other long-term assets. Operating income refers to the income from main business or other businesses. Total profit is the final financial result achieved by the enterprise through production and operation activities in a certain period of time, mainly composed of sales profit and non-operating net income and expenditure.

In terms of basic indicators, indicators such as power consumption of the whole society and regional GDP are selected. Power consumption of the whole society is the total power consumption of all kinds of users in a certain area, including the spontaneous self-use part of the power plant, as well as small-scale photovoltaic power generation, thermal power generation in factories, and gas power generation in coal mines. Regional GDP is the final result of the production activities of all resident units in the region within a certain period of time, equal to the sum of the added value of all industries.

2.2. Correlation analysis method

Use the variable importance judgment function of XGBoost model to analyze the correlation between key indicators. Taking a key indicator as the explained variable, and other key indicators and basic indicators as the explanatory variables. XGBoost model is constructed to judge the importance of the variables based on the coverage of the explanatory variables in the modeling.

XGBoost model is an integrated learning model proposed by Chen et al. (2016)[10] to improve the gradient lifting decision tree model. The decision trees in this model have sequential correlation. Based on the current prediction error of the previous round, the model is constructed iteratively using each round of prediction error to improve the accuracy of prediction.

Assuming that $(x_i, y_i), i = 1, \ldots, n$ is the modeling sample, $\hat{y}_i^{(t)}$ is the prediction result of the model after the t iteration, and $f_t(x_i)$ is the prediction result of the t decision trees, the solution form is:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{1}$$

Since the prediction result of round t-1 is fixed at the time of iteration t, only the prediction function $f_t(x_i)$ needs to be considered when setting the model objective function. The following objective functions should be minimized when solving the model parameters:

$$S^{(t)}(\beta) = L(\beta) + D(f_t) + C \tag{2}$$

Where,

$$L(\beta) = \sum_{i=1}^{n} l(y_i, y_i^{(t-1)} + f_t(x_i)) \tag{3}$$

$$D(f_t) = \gamma T + 0.5\lambda \sum_{j=1}^{T} \omega_j^2 \tag{4}$$

In formula (2), $L(\beta)$ is the loss function of the fitting degree of the measure model, $D(f_t)$ is the regularization term of the complexity of the measure model, and C is the constant term. In formula (3), $l(\cdot)$ is the loss function of the prediction accuracy of the measured sample. In formula (4), T is the number of leaf nodes in the decision tree, ω_j is the prediction result corresponding to the leaf node, and γ and λ are the corresponding adjustment coefficients. By expanding the loss function Taylor to the quadratic term, the greedy algorithm can be used to solve the parameters of the model.

Three indicators of the importance of independent variables will be obtained during the model construction process:

① Cover, which determines the number of leaf nodes of the observed value in each tree when calculating each independent variable as a partition attribute.

② Gain, which calculates the average loss reduction when each independent variable is used as a tree partition attribute.

③ Frequency, which calculates the number of times each independent variable is used to divide attributes in all trees.

The above three independent variable importance measurement indicators are the higher the index value, the higher the independent variable importance. In practice, in order to facilitate the comparison of the importance of independent variables of different sample combinations, the above indicators are often treated with relative quantity.

$$\text{Re}\mathit{lative\ number\ of\ Cover\ of\ independent\ }\text{var}\mathit{iable\ } X_i = \frac{\mathit{Cover\ of\ independent\ }\text{var}\mathit{iable\ } X_i}{\sum_{i=1}^{p} \mathit{Cover\ of\ independent\ }\text{var}\mathit{iable\ } X_i} \tag{5}$$

$$\text{Re}\mathit{lative\ number\ of\ Gain\ of\ independent\ }\text{var}\mathit{iable\ } X_i = \frac{\mathit{Gain\ of\ independent\ }\text{var}\mathit{iable\ } X_i}{\sum_{i=1}^{p} \mathit{Gain\ of\ independent\ }\text{var}\mathit{iable\ } X_i} \tag{6}$$

$$\text{Re}\mathit{lative\ number\ of\ Frequency\ of\ independent\ }\text{var}\mathit{iable\ } X_i = \frac{\mathit{Frequency\ of\ independent\ }\text{var}\mathit{iable\ } X_i}{\sum_{i=1}^{p} \mathit{Frequency\ of\ independent\ }\text{var}\mathit{iable\ } X_i} \tag{7}$$

3. Empirical research

The empirical research data comes from the real production and operation data of a power grid enterprise, which is consolidated from the daily business data sheets of several branches. The sample size of the dataset is 81, and the number of indicators selected is 9.

In terms of production indicators, power supply volume, line loss rate, capacity expansion are selected as the explanatory variables, and other indicators are used as explanatory variables to establish XGBoost model and calculate the relative indicators of coverage as shown in table 1.

Table 1. Coverage of influencing factors of production indicators.

	power supply volume	line loss rate	capacity expansion
power supply volume	/	0.2134	0.4169
line loss rate	0.1208	/	0.1000
capacity expansion	0.3538	0.1515	/
power generation volume	0.0943	0.1368	0.0794
total assets	0.0744	0.1125	0.1273
operating income	0.0624	0.0492	0.0803
total profit	0.0720	0.2137	0.1490
power consumption of the whole society	0.1171	0.0635	0.0117
regional GDP	0.1208	0.0594	0.0355

The key indicators that affect the power supply volume of power grid enterprises are internal management factors such as line loss rate, capacity expansion, and external environmental factors such as power consumption of the whole society and regional GDP. The key indicators that affect the line loss rate of power grid enterprises are internal management factors and operational effectiveness factors, such as power supply volume, capacity expansion, power generation volume, total assets, total profit, etc. The key indicators affecting the capacity expansion of power grid enterprises are power supply volume, total assets, total profit and other factors of operation effectiveness.

In terms of operating indicators, power generation volume, total assets, operating income, total profit are selected as the explanatory variables, and other indicators are used as explanatory variables to establish XGBoost model, and calculate the relative indicators of coverage as shown in table 2.

Table 2. Coverage of influencing factors of operating indicators.

	power generation volume	**total assets**	**operating income**	**total profit**
power supply volume	0.1005	0.1189	0.2939	0.1652
line loss rate	0.1601	0.1991	0.1336	0.2398
capacity expansion	0.1284	0.1213	0.1141	0.0460
power generation volume	/	0.1512	0.1410	0.2317
total assets	0.0469	/	0.0758	0.1601
operating income	0.1919	0.1102	/	0.0677
total profit	0.0813	0.1831	0.1666	/
power consumption of the whole society	0.2423	0.0221	0.0426	0.0442
regional GDP	0.0487	0.0942	0.0323	0.0454

The key indicators that affect the power generation volume of power grid enterprises are production factors such as power supply volume, line loss rate, capacity expansion, as well as operational factors such as operating income and external demand factors such as power consumption of the whole society. The key indicators that affect the total assets are productive factors such as power supply volume, line loss rate, capacity expansion, and operational factors such as power generation volume, operating income and total profit. The key indicators that affect the operating income are the productive factors such as power supply volume, line loss rate, capacity expansion, and the operational factors such as power generation volume and total profit. The key indicators affecting total profit are productive factors such as power supply volume and line loss rate, and operational factors such as power generation volume and total assets.

Based on the above indicator correlation judgment, the following key indicator relationship diagram of production and operation of power grid enterprises can be drawn.As shown in Figure 1.

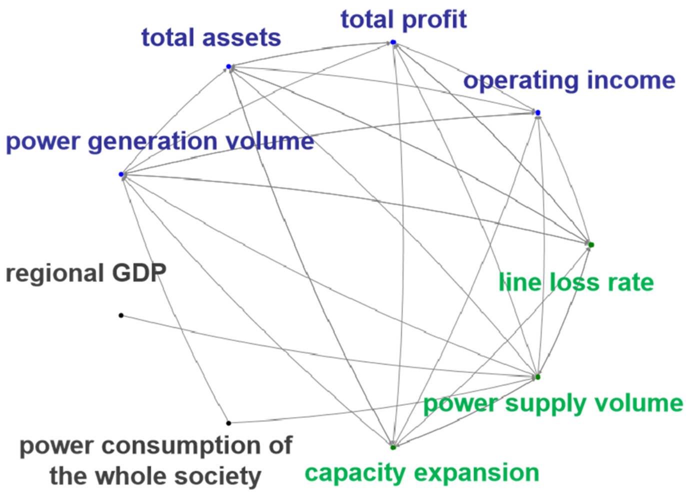

Figure 1. Coverage of influencing factors of business indicators.

4. Conclusion

Based on the variable importance judgment method of XGBoost model, we can judge the strength of the correlation between the key indicators of production and operation of power grid enterprises. The empirical study shows that there is a linkage between the production indicators and the operation indicators of power grid enterprises. Basic indicators mainly affect the power supply or sales of power grid enterprises, and less affect production or operation indicators. Power grid enterprises can better predict and manage production and operation activities according to the relationship between key indicators.

Acknowledgments: This paper is a phased achievement of Science and technology projects of State Grid (Key Technologies for Development Investment and Operation Quality Improvement of the Sgcc's "One Body and Four Wings" under the New Situation, 1400-202157361A-0-0-00).

References

[1] J. Quan, X.P. Ma. Evaluation of enterprise production and operation risks [J]. Industrial Engineering, 1998 (01): 34-37.
[2] J. Wei. Research on the optimization of enterprise business plan management index system [J]. Value Engineering, 2015, 34 (33): 25-28.
[3] P. Zhao. Brief discussion on cost indicators and enterprise production and operation decisions [J]. Mall Modernization, 2015 (Z2): 88-89.
[4] S.G. Zhang. Construction of enterprise cost reduction and efficiency increase management system [J]. Metallurgical Finance and Accounting, 2017 (09): 32-34.
[5] Y.L. Ning. Statistics and analysis of production and operation indicators of thermal power plants [J]. Statistics and Management, 2017 (08): 172-174.

[6] W.Q. Gong. Application of financial indicators and non-financial indicators in enterprise performance evaluation [J]. Finance and Economics, 2019 (20): 143.
[7] C. Zheng, J.C. Li, Y.G. Sun, D.D. Jia, Z.W. Zhao. Exploration and Practice of Multi-dimensional Lean Management Reform of State Grid [J]. Finance and Accounting, 2021 (23): 17-19.
[8] Y.W. Shen. Analysis of production and operation quality of oilfield power enterprises [J]. Modern Industrial Economics and Information Technology, 2022, 12 (04): 219-221.
[9] Y.M. Han, P.F. Xu, J.F. Gong, Y.R. Shen. Research on the comprehensive evaluation system of power grid development and operation based on user-side demand [J]. Power Generation Technology, 2022, 43 (04): 636-644.
[10] T. Chen, C. Guestrin. Xgboost:A Scalable Tree Boosting System[C]. The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016:785-794.