

Research on Investment Trend of Distribution Network Based on Support Vector Machine

Chao Wang¹, Jiyuan Zhang¹, Yaling Jian¹, Jie Jiao¹, Xi Yang¹, Guangxiu Yu¹, Le Du² *

¹ State Grid Sichuan Economic Research Institute, Chengdu, Sichuan, China

² School of Economics and Management, North China Electric Power University, Beijing, China

Abstract: Under the new situation, with the continuous development of my country's economy and the implementation of power system reforms, higher development requirements have been put forward for the distribution network investment plan. Through the scientific and reasonable calculation of the investment scale of the distribution network, optimizing the investment scale of the distribution network and rationally arranging the investment planning of the distribution network project have become one of the key concerns of the current power grid enterprises. This paper uses fishbone diagram theory to analyze the factors that affect the investment scale of the distribution network, and selects the key factor indicators to construct a distribution network investment trend prediction model based on support vector machines. By selecting a certain region's distribution network investment for empirical forecasting analysis, and comparing with the planned investment of the distribution network in the region, the validity of the model is verified.

1 Introduction

As a basic power facility, the distribution network plays an extremely important role in the development of the economy and the improvement of people's living standards. Therefore, strengthening the investment forecast management of the distribution network is of vital importance to improving the operating efficiency of power grid enterprises.

Literature [1] uses five mathematical models to predict the power consumption and maximum load of Wuhai City on the premise of analyzing the development status and problems of Wuhai City to support the investment and construction of the power grid. Literature [2] based on a comprehensive analysis of the current status of distribution network management, combined with historical data, obtained some problems in the distribution network investment planning of county-level power supply enterprises, and then used the gray prediction method and the multiple linear regression prediction model to carry out Load forecasting avoids the shortcomings of a single forecasting method, and achieves the purpose of reducing forecasting risks and increasing forecasting accuracy. Literature [3] built a medium- and long-term investment scale prediction model for provincial power grids based on support vector machines based on analyzing the new environment, new directions, and new requirements of power grid investment. Literature [4] firstly analyzes the main factors affecting power grid investment and its impact mechanism from the macroeconomic and microeconomic aspects, and then builds an ARMA-based power grid

investment prediction model. Literature [5] established a power grid infrastructure investment calculation model using gray theory, and verified the validity and rationality of the model through an example of power grid infrastructure investment calculation in a certain area.

To sum up, the current grid investment forecasts are mostly the application of gray models and linear regression models. Due to the lack of grid-related data and insufficient analysis, the final investment forecast results are relatively one-sided and not accurate enough. Therefore, this paper first applies the fishbone diagram theory to systematically analyze the main factor indicators that affect the distribution network investment; then combines the relevant expert opinions and the measurability of the factor indicator data to select the key factor indicators to construct a support vector machine-based distribution network investment forecast Model; Finally, the relevant data of a certain area distribution network is selected as a sample to verify the validity and feasibility of the model.

2 Analysis of influencing factors of distribution network investment

Under the new situation, the influence factors of power distribution network investment of power grid enterprises are becoming more diversified and complicated. From the perspective of power grid companies, when making distribution network investment planning, in addition to regional economic development plans, industrial development policies, and the company's own

*Corresponding author's e-mail: 1182306191@ncepu.edu.cn

development strategies, it is also necessary to comprehensively consider the optimization of the distribution network structure and the reliable and safe power supply quality. , Electric energy substitution, and promotion of electric energy consumption. This article combines the current situation of distribution network investment planning and management and relevant expert suggestions, from the four dimensions of power demand, grid safety and reliability, promotion of energy consumption, and distribution network technology innovation, combined with fishbone diagram theory, a systematic analysis of the impact on distribution network investment Factors, as shown in Figure 1 below.

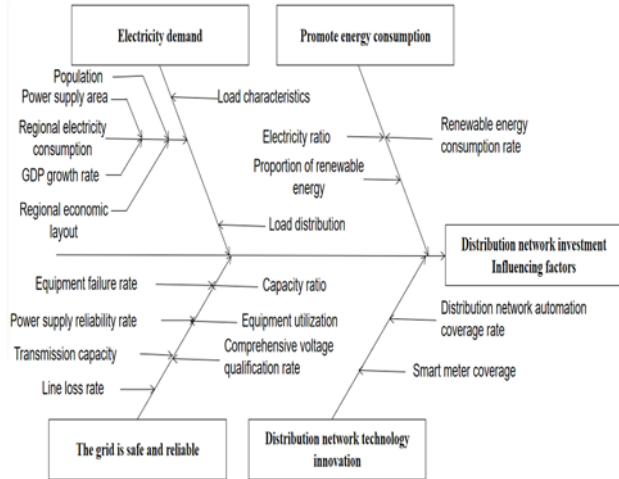


Figure 1. Distribution network investment Influencing factors.

Comprehensively consider the complexity of the forecast model, the quantitative analysis of the influencing factor index data, and the availability of data, combined with the suggestions of relevant experts, select four areas: total electricity consumption, proportion of electricity, reliability of power supply, and coverage of distribution network automation Key indicators are the basic elements of model construction. The basic element index system of the forecasting model is shown in Table 1 below:

Table 1. Predictive model building elements.

Serial number	Indicator type	Index
1	Electricity demand	Regional electricity consumption
2	The grid is safe and reliable	Power supply reliability rate
3	Promote energy consumption	Electricity ratio
4	Distribution network technology innovation	Distribution network automation coverage rate

3 Construction of distribution network investment prediction model based on regression tree theory

The basic idea of the support vector machine is that the function of reducing the dimensional space is transformed into a high-dimensional linear mapping through non-linear mapping, so as to better solve the

problem that the low-dimensional space cannot use linear regression. The basic principles of the period are as follows:

The given sample set is $\{(x_i, y_i), i = 1, 2, \dots, l\}$,

where $x_i \in R^n$, $y_i \in \{+1, -1\}$. The principle of support vector machine for solving nonlinear function optimization is to solve the high-dimensional feature space by linear programming by mapping the nonlinearity to the high-dimensional space. First, we must construct a decision function (classification hyperplane) to classify the sample data. The classification hyperplane can maximize the blank area on both sides of the hyperplane while ensuring the classification accuracy, so as to realize the optimal classification of linear separable problems. Linear separability refers to the use of one or several straight lines to separate sample points belonging to different categories.

If there is a classification hyperplane:

$$WX + b = 0 \tag{1}$$

Meet the constraints:

$$y_i(WX + b) - 1 \geq 0, i = 1, 2, \dots, l \tag{2}$$

It is said that the training set is linearly separable. Where W is the weight vector, X is the input vector, and $b \in R$ is the bias.

Use p to represent the distance between the hyperplane and the nearest sample. The purpose of the support vector machine is to find a hyperplane with the largest separation edge. This hyperplane is the optimal hyperplane (OptimalHyperplane). Therefore, w and b when p is the largest must be determined. The resulting prediction function:

$$Y = f(x) = \text{sgn}(WX + b) \tag{3}$$

Its promotion ability is the best. Among them, $\text{sgn}(\cdot)$ is a symbolic function. That is, the classification error rate of the original sample should be minimized.

From analytic geometry, the distance from any point in the sample space to the optimal hyperplane is:

$$r = \frac{W_0 X + b_0}{\|W_0\|} \tag{4}$$

Simplification:

$$g(x) = r \|W_0\| = W_0 X + b_0 \tag{5}$$

Normalize the function so that $|g(x_0)| = 1$, which is the support vector, is the closest to the classification decision plane, and plays a leading role in the SVM.

Normalize the function so that A , which is the support vector, is the closest to the classification decision plane, and plays a leading role in the SVM. So the algebraic distance from the support vector to the optimal hyperplane is:

$$r = \frac{g(x_0)}{\|W_0\|} = \frac{1}{\|W_0\|} \tag{6}$$

Therefore, it can be considered that maximizing the separation edge is equivalent to minimizing the norm W

of the weight vector. Therefore, the classification hyperplane that minimizes W is the optimal hyperplane.

The classification machine is an algorithm that converts the optimization problem of the maximum interval method to solve the optimal classification surface into its dual problem, so as to solve the original classification problem by solving the relatively simple dual problem.

Using Lagrange optimization method and Wolfe duality theory, the above problem can be transformed into a dual problem, that is, to maximize the functional function:

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$s.t. \begin{cases} \sum_{i=1}^n \alpha_i y_i = 0 \\ \alpha_i \geq 0 \quad i = 1, \dots, n \end{cases} \quad (7)$$

Among them, α_i is the Lagrange multiplier corresponding to sample i .

The optimal classification function obtained by solving the above problems is:

$$f(x) = \text{sgn}(WX + b) = \text{sgn}\left(\sum_{i=1}^{nsv} \alpha_i y_i \langle x_i, x \rangle\right) \quad (8)$$

Among them, nsv is the number of support vectors, and b is the classification threshold, which can be obtained by taking the median of any pair of support vectors in the two categories.

By introducing the relaxation term ξ_i , a generalized classification surface can be realized, and the situation that the training samples are linearly inseparable can be solved, that is, the minimum error classification sample and the maximum classification interval are considered as a compromise. For nonlinear problems, it can be transformed into a linear problem in a high-dimensional space through nonlinear transformation, and then the optimal classification surface is sought in this high-dimensional space. Only the inner product operation $(x_i \cdot x_j)$ between samples is involved, so only the inner product operation needs to be performed in the high-dimensional space, and the inner product operation can be realized by the function in the original space. According to the Hilbert-schmidt principle, as long as the kernel function $K(x_i \cdot x_j)$ satisfies the Mercer condition, it corresponds to the inner product in a certain exchange space.

Therefore, by replacing the inner product in the formula with the kernel function $K(x_i \cdot x_j)$ that satisfies the Mercer condition, a certain linear classification after nonlinear transformation can be realized, and the optimal classification function becomes:

$$f(x) = \text{sgn}\left(\sum_{i=1}^{nsv} \alpha_i y_i K(x_i \cdot x) + b\right) \quad (9)$$

4 Empirical analysis

This paper selects 2016-2019 distribution network investment data in a certain region for simulation analysis. The original data of distribution network investment and electricity consumption levels in the region from 2016 to 2019 are shown in Table 2.

Table 2. Original data table of distribution network investment and electricity consumption in a certain region from 2016 to 2019.

years	Distribution network investment (100 million yuan)	Regional electricity consumption (100 million kWh)	Power supply reliability rate (%)	Electricity ratio (%)	Distribution network automation coverage rate (%)
2016	37.6	365.4	99.936	16.63	25.8
2017	31.2	378.9	99.965	18.98	61.2
2018	28.3	420.7	99.964	25.12	63.7
2019	34.5	543.6	99.968	32.68	88.5

According to the support vector machine algorithm, the predicted value of the distribution network investment scale in the region from 2020 to 2022 in the next three years is as follows:

Table 3. Comparative analysis of model predicted value and planned investment value.

years	Forecast value of distribution network investment (100 million yuan)	Planned investment value (100 million yuan)	Relative error
2020	34.9	35	-0.29%
2021	35.8	35.5	0.85%
2022	35.2	36	-2.22%

As can be seen from the above table, the application model predicts that the investment scale of the distribution network in the region in the next three years will be 3.49 billion yuan, 3.58 billion yuan and 3.52 billion yuan respectively. According to the distribution network investment development plan in the region, the planned investment scale of the distribution network in the region from 2020 to 2022 will be 3.5 billion yuan, 3.55 billion yuan and 3.6 billion yuan, respectively. The relative errors of the investment prediction model are -0.29%, 0.85%, and -2.22% respectively, indicating the model Has good prediction accuracy.

5 Conclusion

In view of the importance of improving the research technology of distribution network investment forecasting, this paper constructs a distribution network investment forecasting model based on support vector machines based on the analysis of the influence factors of distribution network investment, and applies the model to predict 2020-2022 in a certain area The annual distribution network investment scale, through comparison with the planned investment in the region, verifies that the model has good prediction accuracy, and can provide reference and reference for the prediction of the distribution network investment scale of power grid

companies.

Acknowledgments

This research was supported by the Science and Technology Project of the State Grid Sichuan Economic Research Institute "Multi-scale Evaluation of Distribution Network Investment and Comprehensive Optimization Technology Evaluation under the New Reform Situation".

References

1. Wu Rong. Research on Wuhai Distribution Network Planning Scheme[D]. North China Electric Power University, 2018.
2. Liao Junjie. Research on Pingnan County Distribution Network Planning [D]. Guangxi University, 2016.
3. Xie Wei, Pan Wenming, Wang Chenghua, Chen Li, Wang Yang, Ma Jing, Li Zhiwei, Du Haihong. Research on the medium and long-term investment scale prediction model of provincial power grid based on support vector machine[J]. Industrial Technology Economy, 2019, 38(08):154-160.
4. Ji Liwei, Yang Liping, Fei Gaiying. Analysis and prediction of influencing factors of power grid project investment [J]. China Electric Power Enterprise Management, 2016(06): 85-92.
5. Hu Baichu, Hu Gang, Hu Chaohua, Qing Song, Li Mingwei, Peng Chao. Power grid infrastructure investment calculation model based on gray prediction[J]. Journal of University of Electronic Science and Technology of China, 2013, 42(06): 890-894