

Research on Cost Control Line of Overhead Transmission Line Projects in Chongqing

Guangjin Peng^{1,*}, Hui Zhu², Xiaorui Xing², Tingting Feng² and Xuanhong Liang¹

¹Chongqing University(400044), No 174, Shazheng Str, Shapingba District, Chongqing, China

²Chongqing power company, No 21, Zhongshan 3rd road, Yuzhong District (400 16), Chongqing, China

*Corresponding author

Abstract—Investment of Power transmission projects was always huge, and related to many complex factors. Preliminary design of the project in preparation stage plays a very important role in the project cost control. With the application of clustering theory, partial correlation analysis method and support vector machine theory, cost data learning problem of the overhead transmission line projects in Chongqing area was studied in this paper. The results show that the newly presented method to determine cost control line of overhead transmission line projects is scientific and reasonable.

Keywords—overhead transmission line project; cost control line; clustering analysis; partial correlation analysis; support vector machine

I. INTRODUCTION

The investment of Power transmission line projects was huge, and relates to many complex factors, how to control project cost has always been a difficult problem. Preliminary design of the project in preparation stage plays a very important role in the project cost control. According to the current problems of cost control line, Based on the historical data of power project cost, combined with the engineering characteristics of Chongqing area, new methods was developed to further refine the cost control line of overhead transmission line project, which made the cost control line more reasonable and accurate to judge the degree of design optimization and control investment.

II. RELATED TECHNOLOGIES AND THEORIES

A. Clustering Analysis

Clustering analysis technique is a kind of data mining technology, which can separate data objects into many classes or clusters. In the field of power project cost, different projects have difference in project attributes and project costs, it is difficult to find the common characteristics of large number of power projects. By use of clustering algorithm, appropriate project classification can improve this situation, at the same time, through which it is easier to find a suitable and similar project to compare and evaluate project cost of a new project.

In the clustering algorithm of X in the sample space, a data vector is used to represent a sample x (or a feature vector, the observation value), $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$, $x_i \in X, i = 1, 2, \dots, n$, m is sample dimension, n is sample number of X.

Similarity is the basis of the definition of a cluster, and the distance in the feature space is used as a metric to calculate the dissimilarity between the two samples.

$$d(i, j) = \left(\sum_{k=1}^m (x_{ik} - x_{jk})^2 \right)^{0.5} \quad (1)$$

Where, x_{ik} is the observed value of the k- th index of the i-th sample data, x_{jk} is the observed value of the k- th index of the j-th sample data, $d(i, j)$ is the Euclidean distance between the i-th data and the j-th data, much less distance value mean more similar nature between two sample data.

In general, according to the distance measure standard of two sample data, the distance measure standard between two classes, which is a part of clustering algorithm, and also is essential for evaluating the quality of a clustering process.

B. Partial Correlation Analysis

In the correlation analysis, the linear correlation between the two things is realized by calculating the correlation coefficient, and the linear correlation strength between things is determined by the size of the correlation coefficient.

The basic steps of using partial correlation coefficient to analyze the net correlation between variables:

1) Calculate the correlation coefficient of samples. The first order partial correlation coefficient is calculated as follows :

$$r_{y1.2} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{(1-r_{y2}^2)(1-r_{12}^2)}} \quad (2)$$

Where, $r_{y1.2}$ is partial correlation coefficient between y and x1 after excluding the effect of x2. r_{y1}, r_{y2}, r_{12} are correlation coefficients between y and x1, between y and x2, between x1 and x2 respectively.

2) Present original hypothesis. assume there was no significant difference between the partial correlation coefficient and the zero.

3) Construct test statistic. The partial correlation analysis is based on the t statistic, which is used as the test statistic:

$$t = r \sqrt{\frac{n-q-2}{1-r^2}} \quad (3)$$

Where, r is partial correlation coefficient, n is sample number, q is order number. T statistic obeys the t distribution whose freedom degree is $n-q-2$.

4) Calculate the observed value of the test statistic and the corresponding probability value P .

5) Draw the conclusion according to the calculation results.

C. Least Square Support Vector Machine

Least squares support vector machine is a tool which change the solution of the quadratic programming problem into solving linear equations so as to improve the speed and convergence accuracy of the problem

Let X be a $(n+1) \times (n+1)$ matrix, which can be divided into the following form:

$$X = \begin{bmatrix} A & u \\ u^T & a \end{bmatrix} \quad (4)$$

Where, A is a $n \times n$ matrix, u is a $n \times 1$ column vector, a is a scalar quantity. The inverse matrix of the X can be expressed as:

$$X^{-1} = \begin{bmatrix} B & q \\ q^T & t \end{bmatrix} \quad (5)$$

$$X = \begin{bmatrix} A & u \\ u^T & a \end{bmatrix} \quad (6)$$

Where, $B = A^{-1} + tA^{-1}uu^TA^{-1}$, $q = -A^{-1}u$, $t = 1/(a - u^TA^{-1}u)$.
For the least squares support vector machine regression estimation, it is assumed that at $n+1$ moment there is a new training sample (x_{n+1}, y_{n+1}) :

$$\begin{bmatrix} b \\ \lambda \\ \lambda_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1_{1 \times n} & 1 \\ 1_{n \times 1} & \Omega + \frac{I}{\gamma} & p \\ 1 & p^T & \beta \end{bmatrix} \times \begin{bmatrix} 0 \\ Y_{n \times 1} \\ y_{n+1} \end{bmatrix} \quad (7)$$

Where, p is a $n \times 1$ column vector, $p_i = k(x_i, x_{n+1})$, $\beta = k(x_{n+1}, x_{n+1}) + 1/\gamma$, we can get:

$$A_{n+1} = \begin{bmatrix} 0 & 1_{1 \times n} & 1 \\ 1_{n \times 1} & \Omega + \frac{I}{\gamma} & p \\ 1 & p^T & \beta \end{bmatrix} = \begin{bmatrix} A_n & s \\ s^T & \beta \end{bmatrix} \quad (8)$$

In the numerical prediction of LS-SVM, the radial basis kernel function is most commonly used.

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2p^2}\right) \quad (9)$$

III. DATA PREPROCESSING

The actual investment of overhead transmission lines was analyzed. Through collecting relative information of history projects completed and put into operation in 2012 and 2013, original projects cost data was setup.

There was inevitably a lack of uncertainty, inconsistency and redundancy in the actual data collected from the historical projects, so before the establishment of a data mining model the actual data must be processed, this process is called data preprocessing.

Data preprocessing is an important step in data mining and knowledge discovery. Small sample data itself exists some problem such as small number of samples, insufficient information, without data preprocessing, it is more difficult to ensure the accuracy of the learning algorithm.

Data preprocessing was mainly carried out through the following five steps:

- 1) Deletion of unimportant attributes
- 2) Attributes transformation.
- 3) Quantitative treatment of qualitative attributes
- 4) Unit processing of attributes
- 5) Normalization

Here set quantification of qualitative attributes as example to explain the data preprocessing.

In cost data set, the value of attribute "voltage class" was qualitative, such as "220kV" or "110kV", if only unit "kV" was removed, then the value difference between different projects is too large, thus it's necessary to make quantitative treatment.

According to distribution rule of the attribute "voltage class" value, define the following relations to make quantified processing, see table below.

TABLE I. QUANTIFICATION OF ATTRIBUTE "VOLTAGE CLASS"

Before quantification	110kV	220kV
After quantification	1	2

IV. THINNING OF COST CONTROL LINE OF OVERHEAD TRANSMISSION LINE PROJECT

A. Main Influence Attributes Analysis

With the application of clustering algorithm and partial correlation analysis, key attributes which affect project cost obviously were found.

1) Voltage class

By use of clustering algorithm, set cluster class number to 2, iteration times to 200, simulation was made in matlab software, the results of clustering analysis show that the attribute "voltage class" value of most projects were same among all projects in two classes .

Due to expert experience in power cost field, generally there will exist a bit great difference between different voltage projects.

Combined with above-mentioned reasons, the attribute "voltage class" value was selected as key attribute to thin cost control line of overhead transmission line projects.

2) Conductor splitting number

By use of cost data overhead transmission line projects, select the completion time as control variable, make partial correlation analysis of unit length cost and the technical attributes in SPSS software, its results table as follows.

TABLE II. RESULTS OF PARTIAL CORRELATION ANALYSIS

Ctrl variable	coefficient	Unit length cost	Conductor splitting number	Folding length	Loop number	Volatge class	Conductor cross area
Completion time	Unit length cost	1.000	0.434	-0.082	0.541	0.177	0.169
	Conductor splitting number	0.434	1.000	0.074	0.361	0.506	0.426
	Folding length	-0.082	0.074	1.000	0.147	0.261	0.309
	Loop number	0.541	0.361	0.147	1.000	-0.077	0.115
	Volatge class	0.177	0.506	0.261	-0.077	1.000	0.790
	Conductor cross area	0.169	0.426	0.309	0.115	0.790	1.000

From the results table it can be easily found that correlation coefficient between unit length cost and conductor splitting number was the maximum, so in the later study, we selected Conductor splitting number as the second key attribute to thin cost control line

B. Thining of Cost Control Line

By use of two key technical attributes, combining with least squares support vector machine, cost control line of overhead transmission line project was presented.

1) Eliminating of abnormal project cost data

Component heads identify the different components of your pap In order to find out the abnormal information of overhead line project cost data, to avoid the interference of the next step cost control line refinement and determination, these data should be found out and removed.

The fuzzy clustering was applied to analyze the cost data of overhead transmission line project after the pretreatment, and the simulation results are shown in the MATLAB software, just as following figure.

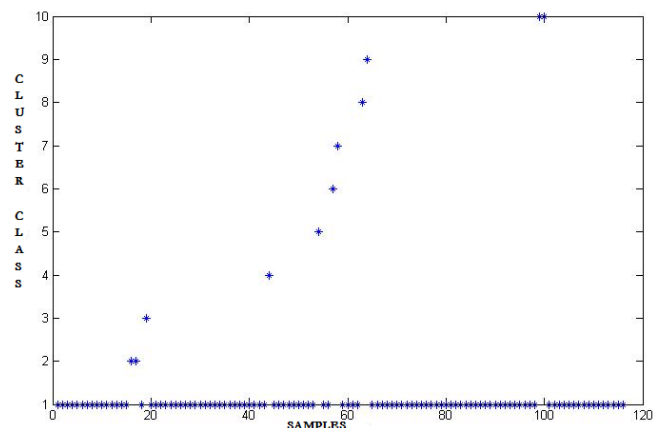


FIGURE I. EXAMPLE OF A FIGURE CAPTION

It can be seen from the above figure that 11 projects didn't belong to first class. These abnormal projects should be removed from the database of historical cost.

2) Refined cost control line

According to the above analysis, using support vector machine regression analysis method for cost data training and learning, final cost control line of 220kV overhead transmission line project was determined as follows.

In order to test the application effect of cost control line, 220kV overhead transmission line projects completed in 2012 and 2013 in Chongqing area were selected to compare the Investment difference ration between national standard ,actual investment and investment determined by refined cost control line in this paper.

TABLE III. REFINED COST CONTROL LINE

Voltage class	Spliting number	Loop number	Cost control line (Yuan/km)	Difference Ratio Compared with national standard (%)	Difference Ratio Compared with actual investment (%)
220kV	4	2	313.23	1.31	-0.26
	2	1	169.36		
	2	1	113.58		
	1	2	195		
	1	1	104.80		

Shown from above table, the refined cost control line was reasonable, highly consistent with the actual investment in Chongqing area.

V. CONCLUSION

In this paper, with the application of artificial intelligence technologies such as support vector machine, particle swarm clustering algorithm and partial correlation analysis, historical cost data of overhead transmission line projects were analyzed to determine the refined cost control line, and the investment determined by the refined cost control line was compared with actual investment and investment determined by national standard, comparison results show that thinning method of cost control line presented in this paper was scientific and reasonable.

REFERENCES

- [1] Shuying Yang. Pattern recognition and Intelligent Computing - Matlab technology realization, 1st ed., Beijing: Electronic Industries Press, 2008, pp.120–168.
- [2] Guangjin Peng, Research on intelligent learning method of small sample data of project cost and its application in power transmission engineering[D].Chongqing:chongqing University, 2010
- [3] Juntao Wei, Study on Small-Sample Estimation Model Used in Power Engineering Cost [D].Chongqing:chongqing University, 2009