Research on Coal Production Cost Prediction Based on PCA-SSA-SVR

Shuntang Zhang, Zhenyang Shi, Lihua Hu and Guojun Zhang
Shandong Technology and Business University, Yantai, China

Keywords: Coal Enterprises, Cost Influencing Factors, PCA-SSA-SVR, Cost Prediction Model.

Abstract: This paper starts from the research perspective of lean market-oriented management mechanism within coal enterprises, establishes key influencing factors indicators in terms of environment, technical equipment and organizational management, and builds a cost prediction model based on PCA-SSA-SVR, and compares it with multiple regression analysis and PCA-BP prediction model, the results show that the proposed model has outstanding It has the advantages of avoiding dimensional disasters, overcoming the shortcomings of relying on empirical debugging penalty coefficients and kernel function parameters, and the prediction accuracy meets the requirements, which can provide a basis for modern coal enterprises to formulate labour quotas and cost control plans.

1 INTRODUCTION

The coal industry plays a crucial role in China's economic development and energy supply (HOU Xiaochao, 2020). However, in the current economic and energy structure transformation period, the coal industry is facing the challenge of slowing coal demand (LIU Chang, 2017). To enhance their competitiveness, coal enterprises must focus on improving their cost advantage, making refined production cost management increasingly important (XU Bo, 2013). Compared to traditional cost management methods, the internal lean market-based management approach incorporates lean thinking and a market-oriented perspective. It breaks down costs into specific tasks and processes, optimizes resource allocation within the organization, and enables finer control over enterprise costs (JIANG Zhonghui, 2018). The introduction of mechanized and intelligent equipment has also brought changes to the cost structure of coal production (LI Guoqing, 2022). To optimize the existing cost management system in China's coal enterprises, it is crucial to conduct a comprehensive analysis of production factors, design a production cost forecasting index system, scientifically forecast coal production cost trends, and reduce subjectivity in decision-making.

Zhiling and Jiahao (REN Zhiling, 2015) developed a scheduling scheme based on a grey prediction mathematical model to predict the relationship between water inflow and time in roadways. This scheme effectively reduces electricity costs. Hossain (Hossain M E, 2015) and Meng (ZHU Meng, 2015) introduced a dynamic approach to cost analysis by studying the factors that influence costs, departing from the traditional static research method. Data mining and intelligent algorithms have gained popularity in recent years, leading researchers to explore both new methods and existing research results for potential improvements. Noural (Nourali H, 2018) used a support vector regression machine prediction model for cost estimation. Jing (YANG Jing, 2017) et al. improved the prediction accuracy of coal logistics cost by constructing a support vector regression machine based on the chicken swarm optimization algorithm. Xiaohong and Huijia(TAI Xiaohong, 2017) enhanced the prediction accuracy of coal mining cost by utilizing an improved adaptive particle swarm optimization algorithm to determine the penalty factor and kernel function for the least squares support vector machine, resulting in satisfactory outcomes.

In summary, the industry has developed a comprehensive set of cost forecasting ideas and continuously improved forecasting methods. However, the changing mining mode in modern coal mining enterprises has altered the structure of production cost elements. Additionally, existing literature lacks sufficient focus on the current state

Zhang, S., Shi, Z., Hu, L. and Zhang, G.
Research on Coal Production Cost Prediction Based on PCA-SSA-SVR.
DOI: 10.5220/0012286300003807

In *Proceedings of the 2nd International Seminar on Artificial Intelligence, Networking and Information Technology (ANIT 2023)*, pages 483-487
ISBN: 978-989-758-677-4

of coal enterprise cost management, mainly due to a lack of firsthand data. Furthermore, comprehensive research progress has revealed that cost forecasting models in coal enterprises heavily rely on historical cost data, but the processing of acquired data before forecasting has been neglected.

2 ANALYSIS OF PRODUCTION COSTS FOR COAL COMPANIES

2.1 Coal Enterprise Mining Characteristics

At present, with the increasing amount of mechanical and electrical equipment invested in coal enterprises, maintenance and depreciation costs are gradually increasing, while the complex mining environment due to the limitations of special natural factors has led to changes in the components of production costs, making it difficult to accurately forecast production costs for coal enterprises.

(1) Equipment upgrade. Coordinated interaction between technology and management innovation in coal enterprises, mechanised equipment gradually replacing manual operations, labour productivity and coal production increased significantly, but the high introduction cost of large equipment, depreciation and maintenance costs increased cost pressures.

(2) Coal production operations are complex. Unlike manufacturing enterprises, coal production does not consume raw materials and auxiliary materials do not constitute the product entity, which also leads to a complex cost composition and increases the costing workload (ZHANG Qing, 2000).

(3) The operating environment is influenced by natural factors. The production process of coal is limited by natural conditions such as geological formations, reserves and ambient temperature, and has high auxiliary costs such as safety.

2.2 Internal Lean Market-Based Cost Management Mechanisms

The internal lean market-based management framework, depicted in Figure 1, introduces a shift from traditional administrative subordination within coal enterprises. It empowers the enterprise, district teams, teams, and individual positions to operate as market entities. This decentralization of management authority creates trading markets between the various levels of entities. Based on historical operational data, the framework determines the fixed unit price for the trading markets at each level, taking into account the existing economic, technical, and production levels. This approach establishes a new cost management mechanism that incorporates the value chain. By fully engaging the employees and enhancing the enterprise's fine management capabilities, the framework aims to reduce costs, increase efficiency, and enhance competitiveness. It is important to consider the influence of the management level of coal enterprises and market quotas on costs when making cost forecasts within this framework.

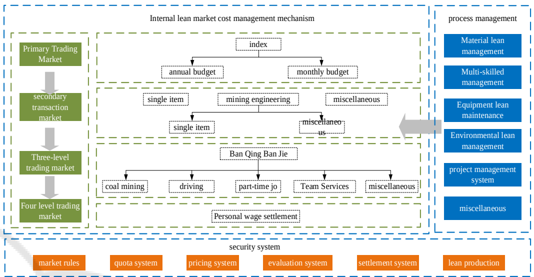

Figure 1: Internal lean market management mechanism.

2.3 Building a Framework of Factors Influencing Coal Production Costs

Through Web of Science, China Knowledge Network and other databases, we have collected and screened the literature that has profound research on coal enterprises' production cost influencing factors and modelling prediction and control in the past 15 years, combined with the current operating environment, internal lean market-oriented management situation and production factor structure of coal enterprises for analysis and integration, finally identified and screened out coal production cost influencing factors and constructed coal The framework of production cost influencing factors is shown in Table 1.

Table 1: Influencing factors of coal production cost.

Influencing factors of coal production cost	
environmental factor	Operating temperature, height mining, Thickness of caving roof, inflow of water
Technical equipment factors	Unit footage output, Single cycle yield, Lifting efficiency, Equipment service life, mechanisation level, Mining coordination level
the lean management level	standardization level, Proposal improvement rate, safety assessment, labour quality, quality of manager
Marketized quota management	Material quota, manual unit price, electricity expense, maintenance cost unit price, and
others	coal quality, supplying difficulty

3 COST FORECASTING MODEL BASED ON PCA-SSA-SVR

3.1 Principal Component Analysis

Let the coal production cost influencing factors be $X_1, X_2, \cdots, X_n$, create a correlation type matrix X for $X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nn} \end{bmatrix}$. Arrange the non-negative eigenvalues of the correlation coefficient matrix λ_i in the order of $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq 0$, and according to $\alpha_i = \lambda_i / \sum_{i=1}^{n} \lambda_i, i = 1, 2, \cdots, n$, count the cumulative contribution of the former as p $\alpha = \sum_{i=1}^{p} \alpha_i$. When $\alpha \geq 85\%$ is selected, the eigenvector $r_1, r_2, \cdots, r_p$ corresponding to the eigenvalue $\lambda_1, \lambda_2, \cdots, \lambda_p$ is selected to construct the principal component matrix. $R_j = \left[r_{j1}, r_{j2}, \cdots, r_{jn} \right]^T$ $(j \leq p)$

The principal components were extracted from the original data at p and the combined principal component score was calculated using the following formula

$$FACj = r_{j1} \cdot X_1 * + r_{j2} \cdot X_2 * + \cdots + r_{jn} \cdot X_n *, (j = 1, 2, \cdots, p)$$

3.2 Sparrow Search Algorithm

The Sparrow Search Algorithm (SSA) is inspired by the sparrow's ability to complete foraging and anti-predatory behaviour by updating the positions of finders, followers and vigilantes, and has the advantages of strong global search capability and fast convergence(Zhang K, 2023).

Discoverer location update formula in SSA is

$$X_{i,j}^{e+1} = \begin{cases} X_{i,j}^{e} \dfrac{dy}{dx} \exp\left[-i / (\alpha D) \right] & if (R_2 \in (0, S)); \\ X_{i,j}^{e} + QL & if (R_2 \in \left[S, 1)\right), \end{cases}$$

In the above equation, $j = 1, 2, \cdots, d$ and d are the population dimensions; e is the current number of iterations; D is the maximum number of iterations; Q is a random number with a normal distribution; X_{ij} is the location of the sparrow i in the j dimension; $\alpha \in (0,1]$ is a uniform random number; $R_2 (R_2 \in [0,1])$ and $S (S \in [0,1])$ are the warning and safety values, respectively; L is a $1 \times d$ matrix with each element being 1.

Accession position update formula:

$$X_{i,j}^{e+1} = \begin{cases} Q \exp\left[(X_{worst} - X_{i,j}^{e}) / i^2 \right] & if (i \in (m/2, +\infty)); \\ X_{p}^{e+1} + \left| X_{i,j}^{e} - X_{p}^{e+1} \right| \cdot A^{+} \cdot L & if (i \in (m/2, +\infty)), \end{cases}$$

In the above equation: X_{worst} is the current global worst position; X_p is the best position occupied by the current finder; A is a matrix of $1 \times d$ with -1 or 1 elements each and $A^{+} = A^T (AA^T)^{-1}$; the number of sparrows is m , when $i > m/2$ indicates that the first i joiner is less adapted and not getting food and needs to fly to another position to feed for energy.

When a hazard is detected, the vigilante position is updated with the formula:

$$X_{i,j}^{e+1} = \begin{cases} X_{best}^{e} + \beta \left| X_{i,j}^{e} - X_{best}^{e} \right| & if (f_i \in (f_g, +\infty)); \\ X_{i,j}^{e} + \dfrac{K \left| X_{i,j}^{e} - X_{worst}^{e} \right|}{(f_i - f_w) + \varepsilon} & if (f_i = f_g), \end{cases}$$

In the above equation: ε is the minimum constant; $K \in [-1,1]$ is a random number; β is a random number that follows the mean standard normal distribution, X_{best} is the global current best position; f_i is the sparrow's fitness value; f_g and f_w are the global best and worst fitness values.

3.3 Support Vector Regression Models

Support vector regression model (SVR), a derivative branch of support vector machine (SVM), introduces a fitted loss function to solve the regression problem for non-linear systems. SVR is not only capable of separating input vectors in a multi-dimensional space with a maximum distance hyperplane, but it also has better results in predicting small sample data(Zhou Z, 2022).

The SVR function can be represented as $f(x) = w^T \varphi(x) + b$ where $\varphi(x)$ is a non-linear function, $f(x)$ represents the predicted output and w and b are the corresponding coefficients.

Also the SVR is an optimisation problem and can be expressed as $\min\limits_{w,b,\xi,\xi^*} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n} (\xi + \xi^*)$. Where n is the sample size, ξ and ξ^* are the relaxation variables and $C > 0$ is the regularisation factor.

Also the dual form of the optimisation problem can be obtained using the Lagrangian equation which can be represented by the equation.

$$\max_{\alpha_i,\alpha_i^*} \sum_{i=1}^{n} y_i(\alpha_i-\alpha_i^*)-\varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)-\frac{1}{2}\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)k(x_i,x_j)$$

where $\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0$, $0\le\alpha_i$, $\alpha_i^*\le C$, and the SVR function can be expressed as $f(x)=\sum_{i=1}^{n}(a_i-a_i^*)k(x_i,x_j)+b$. Where α_i and α_i^* are Lagrange multipliers; $k(x_i,x_j)$ is the kernel function and the expression for the Gaussian RBF kernel function chosen for this study is defined as $k(x,x_i)=\exp(-\gamma\|x-x_i\|^2)$, $\gamma>0$.

3.4 PCA-SSA-SVR Prediction Model Construction

In this paper, a PCA-SSA-SVR model is established to forecast coal production costs, and its specific steps are shown in Figure 2.

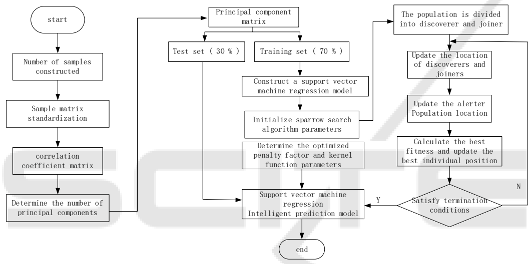

Figure 2: PCA-SSA-SVR model process.

4 MODEL APPLICATIONS

4.1 Data Collection

Using a coal mine in Shaanxi as an example to validate the proposed prediction model, the production information of the mine for the past three years was collected, including market-based labour quota standards, lean production performance assessment data, each team's staffing table, material consumption data, equipment replacement and maintenance records, production per shift, internal lean market-based accounting data of the mine, underground operating environment, operating workers, production equipment and other information, and after sorting and filtering steps to The data was aggregated to form the set containing 21 coal production cost impact factors.

4.2 Analysis of Prediction Results

The PCA method was used to reduce the dimensionality of the 21 influencing factors indicators of coal production cost, and the calculated principal component analysis data was used as the input data of the prediction model, corresponding to the marketed unit cost per tonne of coal of that district team as the output layer. The population size was set to 30 during the model training, the maximum number of iterations was 100, and the penalty coefficient and kernel function parameters were taken to be between 0.001 and 1000.

The optimal penalty coefficient and kernel function parameters after algorithm optimization are 388.705 and 113.337 respectively. The above optimal parameters are brought into SVR. The training results of PCA-SSA-SVR prediction model show that the correlation coefficient between the actual value and the predicted value is $R>0.88$, indicating that the hybrid model has strong learning ability and high prediction accuracy. The remaining 12 sets of data are used as prediction samples. After training various prediction models, the prediction results are shown in Figure 3. It can be seen intuitively that the error between the predicted value and the actual value of the PCA-SSA-SVM model is the smallest, and it has better prediction ability.

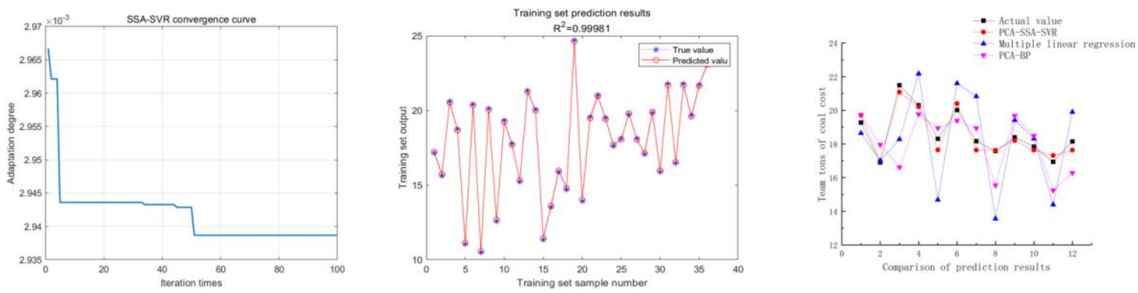

Figure 3: Training results of PCA-SSA-SVR prediction model.

5 CONCLUSION

Aiming at the problem of cost forecasting in modern coal enterprises, this study screens and summarizes the key influencing factors of coal production costs based on the internal lean market management mechanism of coal enterprises, establishes the PCA-SSA-SVR cost forecasting model, and applies and verifies the model. Finally, the following three conclusions are drawn：

(1) Compared with the traditional method of distributing the cost of machinery and equipment according to the standard, the internal lean market-oriented management mechanism implements the lean improvement system within the enterprise, introduces a market-oriented mechanism, and formulates labor quotas, which is more conducive to the realization of fine-grained enterprise costs management.

(2) By analyzing and summarizing the interrelationship and change law between coal production cost and various influencing factors, this study comprehensively establishes the key factor indicators that affect coal production cost from the aspects of environment, management level, and marketization quota, so as to ensure that coal production cost Scientific Validity of Predictions.

(3) The results of the model application test show that: based on the PCA-SSA-SVR model, the efficient and accurate prediction of production costs can be realized, which can provide a basis for coal enterprises and other fields to formulate labor quotas and cost control plans, and has certain promotion and application for coal enterprises value.

REFERENCES

HOU Xiaochao, ZHANG Lei, YANG Qing. Chinese medium and long-term coal demand forecast based on Monte Carlo method (J). Operations Research and Management Science, 2020, 29(01): 99-105.

LIU Chang, SUN Chao. Long-term future prediction of China's coal demand (J). China Coal, 2017, 43(10): 5-9+20.

XU Bo, HUANG Wusheng. Management system of accurate ore actual valueactual valuecosting for gold mines (J). Metal Mine, 2013(05): 125-127.

JIANG Zhonghui, LUO Junmei, MENG Chaoyue. A study on the path of dual-perception to overcoming inertia (J). Journal of Zhejiang University (Humanities and Social Sciences), 2018, 48(06): 171-188.

LI Guoqing, WU Bingshu, HOU Jie, et al. Mining cost prediction model for underground metal mines (J). Metal Mine, 2022(05): 62-69.

REN Zhiling, HAN Jiahao. Energy saving control research on mine drainage system based on model predictive control (J). Journal of System Simulation, 2015, 27(12): 3032-3036+3043.

Hossain M E. Drilling costs estimation for hydrocarbon wells (J). Journal of Sustainable Energy Engineering, 2015, 3(1): 3-32. https://doi.org/10.7569/jsee.2014.629520.

ZHU Meng. Study on safety cost of coal companies based on systematic dynamics model (J). Coal Engineering, 2015, 3(1): 3-32.

Nourali H. Mining capital cost estimation using Support Vector Regression (SVR) (J). Resources Policy, 2019, 62: 527-540. https://doi.org/10.1016/j.resourpol.2018.10.008.

YANG Jing, LI Junfu, ZHANG Gaoqing. Coal logistics cost prediction based on improved support vector regression (J). Journal of Guangxi University (Natural Science Edition), 2017, 42(04): 1623-1627.

TAI Xiaohong, ZHANG Huijia. Coal mining cost prediction model based on IAPSO-LSSVM (J). Journal of Liaoning Technical University (Natural Science), 2017, 36(05): 554-560.

ZHANG Qing, WANG Dongping, LI Xingdong. An analysis on main factors affecting cost of enterprise of coal mine (J). Journal of Liaoning Technical University (Natural Science), 2000(03): 323-326.

Zhang K, Chen Z, Yang L, et al. Principal component analysis (PCA) based sparrow search algorithm (SSA) for optimal learning vector quantized (LVQ) neural network for mechanical fault diagnosis of high voltage circuit breakers(J). Energy Reports, 2023, 9: 954-962. https://doi.org/10.1016/j.egyr.2022.11.118.

Zhou Z, Dai Y, Wang G, et al. Thermal displacement prediction model of SVR high-speed motorized spindle based on SA-PSO optimization (J). Case Studies in Thermal Engineering, 2022, 40: 102551. https://doi.org/10.1016/j.csite.2022.102551.