Cost Estimation of Manufacturing Enterprises based on BP Neural Network and Big Data Analysis

HUIJUAN MA
Lyceum of the Philippines University Manila Campus,
Manila 1002,
PHILIPPINES

Abstract: **-** The manufacturing industry is the pillar industry of modern industry, and the cost estimation of manufacturing enterprises is an important management means of the manufacturing industry. Aiming at the cost estimation problem of manufacturing enterprises, this research proposes a cost estimation method based on Back Propagation (Back Propagation) neural network and big data analysis. In the process, the Lambda architecture was used to construct the big data analysis architecture of manufacturing enterprises, the K-means clustering algorithm was introduced for data clustering, and then the genetic algorithm was combined with the Back Propagation neural network to estimate the cost. In the estimation accuracy test, the accuracy of the research method can reach 94.7% after 240 iterations; in the calculation time test, the calculation time of the research method is 403 Ks when the data size is 500 Gb in a large-scale data set; in the call data volume test, the call data volume of the research method is 164 Kb when the research method is carried out to the seventh step in the small-scale data set; when the application analysis is carried out, the research method completes accurate cost estimation for 9 target parts. This research method has good model performance and calculation accuracy, and can effectively estimate manufacturing enterprises' costs.

Key-Words: **-** Back Propagation; Lambda architecture; Big data; Cost estimation; Manufacturing; K-means.

Received: April 15, 2023. Revised: October 14, 2023. Accepted: November 6, 2023. Published: November 17, 2023.

1 Introduction

In the manufacturing industry, cost estimation is a crucial part of enterprise decision-making and business management. Accurate cost estimation can help enterprises formulate reasonable pricing strategies, optimize production processes, and improve profit margins and competitiveness, [1]. Traditional cost estimation methods usually require a large amount of data collection and processing, and these data are often scattered, inconsistent, or missing, which brings difficulties to cost estimation, [2], [3]. The models in traditional cost estimation methods are usually based on simplified assumptions and empirical formulas, ignoring the complex manufacturing environment and the interaction between multiple influencing factors, resulting in limited accuracy of estimation results, [4]. A back propagation neural network (BPNN) is a commonly used artificial neural network model. By learning training data, it can discover the nonlinear relationship between input features and costs, and can automatically adjust the relationship between neurons through the backpropagation algorithm. Connection weights between them, so as to achieve accurate cost estimation, [5]. Big data technology can help manufacturing companies process and analyze massive amounts of data, thereby providing an accurate basis for cost estimation. [6], proposed a cost estimation method for enterprises that combines building information models with target value design. This method can analyze risks and profits, but its computational efficiency at runtime is relatively average. Scholars such as Mishra S have designed an enterprise cost evaluation method using an ant colony algorithm and resource joint allocation algorithm. This method can propose optimization strategies for costs, but the accuracy of cost estimation is relatively average, [7]. In view of this, research attempts to design a method based on the BP neural network to analyze big data and obtain cost estimation results for manufacturing enterprises. Utilize BP neural network and genetic algorithm to achieve efficient cost estimation for manufacturing enterprises, and additionally introduce K-means clustering algorithm to improve the computational accuracy of the research method. The aim is to integrate the advantages of various technological means to design a cost estimation method for manufacturing enterprises with better performance, providing feasible technical references for the development of manufacturing enterprises.

The research is mainly carried out in four parts. The first part discusses and summarizes the relevant research results of the current cost estimation and

BPNN. The second part is mainly to design the cost estimation method for manufacturing enterprises based on BPNN and big data analysis. The third part is the performance test and empirical analysis of the research method. The last part is the discussion and summary of the full text.

2 Related works

Cost estimation of manufacturing enterprises can provide a reference for production planning and development of enterprises, and many scholars have conducted related research on cost estimation methods. Scholars such as Fazeli proposed a BIM-based estimation method for the cost estimation of construction projects. In the process, the material quantity calculation is associated with the model and expanded. This proposed method can effectively estimate the construction cost, [8]. Scholars such as Nevliudov proposed an estimation method using a regression model for the estimation of material cost in 3D printing. In the process, the resin consumption and exposure parameters in printing are correlated, and the correlation coefficient of the circuit board topology is calculated. This proposed method has high computational accuracy, [9]. [10], proposed a method based on multi-factor analysis for the cost estimation of automatic mobile phone washing in public places. During the process, the usage habits of the hand-washing crowd are analyzed, and the raw materials are combined for calculation. This proposed method shows high accuracy. Scholars such as Leelathanapipat proposed a method based on multiple linear regression for the cost estimation of equipment renovation. In the process, three data related to maintenance were introduced as independent variables, and the decision coefficient of the model was adjusted. This proposed method can effectively estimate, [11]. Scholars such as Rosa proposed an estimation model based on scale measurement for the cost estimation of agile software. In the process, the workload of the project is provided, and the application domain group is introduced to improve the accuracy. Experimental results show that the proposed method has good performance, [12].

Some scholars have conducted related research on BPNN. [13], proposed a method based on BPNN for the performance prediction of solid oxide fuel cells. In the process, the support vector machine and the random forest technology are integrated, and the model evaluation is performed using multiple criteria. Experimental results show that the proposed method has good prediction accuracy. Scholars such as Li proposed a prediction method based on BPNN for the early warning of financial risks in business operations. In the process, the initial financial problems are analyzed, and the model is reasoned in the process. This proposed method has a high prediction accuracy, [14]. Aiming at the problem of real-time traffic monitoring of roads, scholars such as Liu proposed a data monitoring system based on BPNN. In the process, the floating car data is fused with the fixed detector data, and the genetic algorithm and ant colony algorithm are introduced to improve the calculation accuracy of the model. This proposed method is effective for condition monitoring, [15]. For the diagnosis of lung cancer, scholars such as Nanglia proposed a detection image analysis method based on BPNN. In the process, the support vector machine is used to simplify the computational complexity, and the feed-forward and back-propagation neural network is integrated to strengthen the features. This proposed method has good diagnostic accuracy, [16]. [17], proposed an analysis method based on BPNN for the data collection of physical components in waste products. The process combines the physical composition of solid waste with social factors and uses hyper-spherical changes to remove constraints. This proposed method has good data analysis performance.

To sum up, although BPNN has been researched and applied in many fields, there is still little research on the cost estimation of manufacturing enterprises. In view of this, the study proposes a manufacturing enterprise cost estimation method based on BPNN and big data analysis, to provide more references for the field of manufacturing cost estimation.

3 Design of Manufacturing Enterprise Cost Estimation method based on BPNN and Big Data Analysis

An effective cost estimation method can obtain accurate product manufacturing cost analysis results. This section will describe the technical means used in the manufacturing enterprise cost estimation method of the research design.

3.1 Big Data Analysis Architecture of Manufacturing Enterprises based on Lambda Architecture

Under the trend of industrial manufacturing intelligence transformation, industrial big data has become an important information carrier in the industrial field, and it collects manufacturing information from a comprehensive perspective, [18], [19].

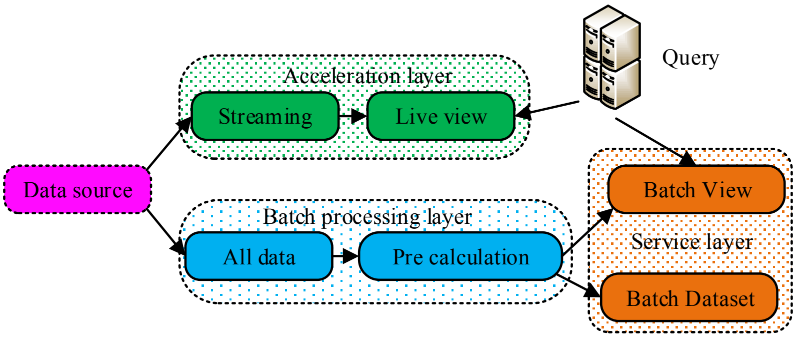

Fig. 1: Lambda architecture

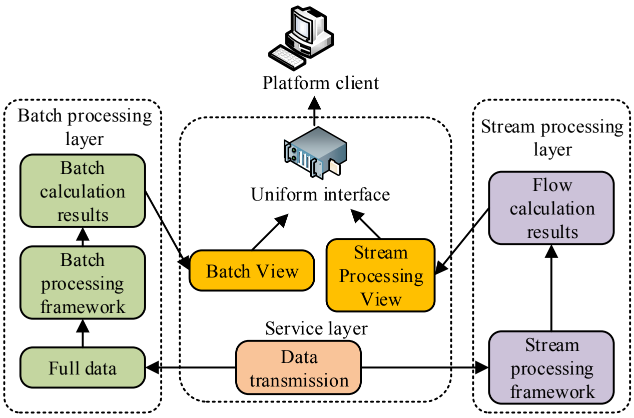

Fig. 2: Multi-mode big data processing architecture

In the application of big data, the use of data for analysis is the most important link. The Lambda architecture can perform real-time stream processing and can be applied to the processing of large-scale complex data, [20]. The research uses the Lambda data processing architecture as the foundation to construct the big data analysis architecture of manufacturing enterprises. The Lambda architecture is shown in Figure 1.

As can be seen from Figure 1, the Lambda architecture includes three main parts: the acceleration layer, the batch processing layer, and the service layer. The data source is connected to the acceleration layer and the batch processing layer, and the query information is connected to the acceleration layer and the service layer. The batch processing layer can perform batch processing calculations, generate batch processing views, and transfer data to the service layer for storage. The batch processing layer can be repeated periodically when generating batch views to improve data fault tolerance and is suitable for computing and analysis on a global scale. The service layer provides support for the input of query information, accesses the view with the query conditions contained in the query information, and calls the real-time view combined with the batch processing results to give feedback to the user. Generally, to ensure the simplicity of the overall system, it is not allowed to Random writes are performed in the result. The acceleration layer processes only the latest input data to reduce processing latency while completing real-time view generation. The data of manufacturing enterprises involves many aspects and has many data models. The Lambda architecture is optimized to establish a multi-mode manufacturing enterprise big data processing architecture. The multi-mode big data processing architecture is shown in Figure 2. From Figure 2 the multi-mode big data processing architecture contains three main layers: batch processing layer, service layer, and stream processing layer. The service layer is the only medium connecting the batch processing layer, stream processing layer, and platform client, providing the same access interface. During data transmission, static historical data is directly imported by the client, and real-time data is processed by Kafka and the distributed coordinator. The batch processing layer uses the batch processing framework to perform offline calculations on the full amount of industrial data in the distributed database system and outputs the calculation results to the batch view. The historical results are not retained during calculations to ensure the timeliness of output data. The stream processing layer uses the stream processing framework to process multiple real-time stream data online, and at the same time sends the latest calculation results to the stream processing view. The batch processing framework uses the Hadoop framework, which includes a distributed storage layer, a resource scheduling layer, and a batch processing engine. The distributed storage layer can store and replicate cluster nodes and store results; the resource scheduling layer manages and schedules basic resources; and the batch processing engine for data calculation. The stream processing framework uses the Storm framework and uses two modes for the combined operation to perform strict one-time processing on the received data. To analyze the data association in the data, the study introduces the K-means clustering algorithm for data clustering. K-means divides the cluster center and calculates the distance between the data and the cluster center to divide and update the cluster to achieve the clustering of the data set. When calculating, the number of input data is first determined, and then the initial dataset is set to specify multiple initial clustering centers. Cluster the data using Euclidean distance, and then divide the unpartitioned data into clusters with the same number of initial cluster centers, as shown in formula (1).

$$dist(x,C) = \min_{x \in X, c \in C} \|x-c\| \tag{1}$$

Formula (1), x represents the data point; C represents the cluster; c and represents the cluster center. After clustering the data, update the clustering center with the clustering results, as shown in formula (2).

$$c_i = \frac{1}{|NC_i|} \sum_{x_i \in c_i} x_i \tag{2}$$

Formula (2), c_i represents the new cluster center of the cluster; NC_i represents the number of data points in the cluster. Set a termination condition and calculate the sum of squared errors. If the sum of squared errors is less than the initial threshold, stop the iteration. The error sum of squares is calculated as shown in formula (3).

$$E = \sum_{i=1}^{k} \sum_{p \in c_i} dist(x,C)^2 \tag{3}$$

Formula (3), E represents the sum of squared errors. Introducing parallel coordinates for dimensionality reduction visualization of multidimensional data. In a multidimensional space, if the dataset contains multiple data and each data contains multiple field attributes, the definition of a parallel dataset is shown in formula (4).

$$D = \left\{ D_m = \left(D_{m,1}, D_{m,2}, \ldots, D_{m,n}, \ldots \right) \middle| 1 \le m \le M, 1 \le n \le N \right\} \tag{4}$$

Formula (4), M it represents the maximum number of data contained; N it represents the maximum value of the field attribute of the data. Calculate the relative position value of each data in the coordinates, as shown in formula (5).

$$p_n = \frac{D_{m,n} - \min D_n}{\max D_n - \min D_n} \tag{5}$$

Formula (5), p_n represents the relative position value. Draw data in parallel coordinate systems using relative position values. The clustering accuracy analysis formula in the follow-up accuracy analysis is shown in formula (6).

$$A = 1 - \left(mis / n_y \right) \tag{6}$$

Formula (6), A represents the accuracy rate; mis represents the number of misclassified samples; n_y represents the total number of samples. The calculation formula for subsequent speedup ratio analysis is shown in formula (7).

$$S_{speedup} = \frac{T_s}{T_r} \tag{7}$$

Formula (7), $S_{speedup}$ represents the speedup ratio; T_s represents the serial execution time on a single node; T_r and represents r the parallel execution time of a computing node.

3.2 Design of Cost Estimation Method based on Improved BPNN

When carrying out cost estimation based on the big data of manufacturing enterprises, due to the different influence of parameters involved in different products, there are defects in the characteristics of the calculation, [21], [22]. As a kind of artificial neural network, BPNN has strong adaptability when estimating product cost. The study uses BPNN to estimate costs based on big data of manufacturing enterprises. The BPNN model is shown in Figure 3.

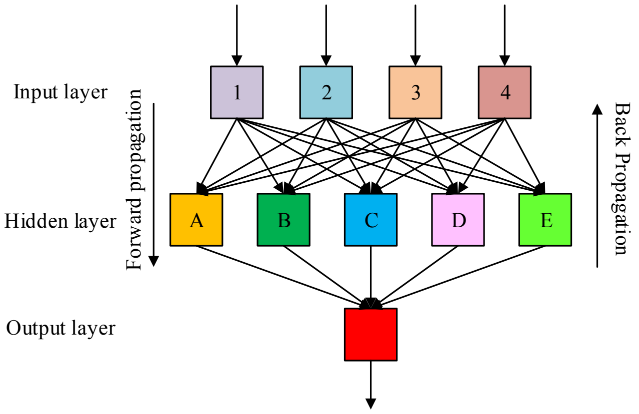

Fig. 3: BPNN model

It can be seen from Figure 3 that the BPNN contains an input layer, a hidden layer, and an output layer. The input layer is the layer that accepts data input, can normalize the data, and performs data buffering at the same time. The hidden layer is set by user requirements, and the number of layers is determined by repeated combinations of data. If there is a large discrepancy in the output of the output layer, the network stops outputting and enters the backpropagation process, corrects the attribute values, and then returns to the forward propagation. In the BPNN, the relationship between neurons is described by the activation function. During the training process, the error, threshold, and weight need to be reduced until they are less than the present value. In the process of information forward propagation, the output calculation of a neuron in the middle layer is shown in the formula (8).

$$z_m = f_1\left(\text{int}_m\right), m = 1,2,...,q \tag{8}$$

Formula (8), z_m represents the output of a neuron in the middle layer; int_m represents the information transfer from the input layer to the output layer; f represents the activation function. The calculation of information transfer is shown in formula (9).

$$\text{int}_m = \sum_{i=1}^{n} v_{im} x_i \tag{9}$$

Formula (9), v_{im} represents the input layer information. The output of the last layer of neurons is shown in formula (10).

$$o_n = f_2\left(\sum_{k=1}^{e} w_{mn} z_m\right), n = 1,2,...,q \tag{10}$$

Formula (10), o_n represents the neuron output of the last layer; w_{mn} represents the information of the middle layer; z_m and represents the information of the last layer. The root mean square error calculation of the forward pass is shown in the formula (11).

$$E_z = \frac{1}{2}\left[\sum_{n=1}^{l}\left(y_n - o_n\right)\right]^2 \tag{11}$$

Formula (11), E_z represents the root mean square error of forward transmission; y_n which represents the real output value of the last layer. In the process of information backpropagation, the weight is adjusted by solving the partial derivative, and the root mean square error of the forward transmission is expanded, as shown in the formula (12).

$$E_z = \frac{1}{2}\sum_{n=1}^{l}\left\{\left[y_n - f_2\left(\sum_{n=1}^{n} w_{im} z_m\right)\right]\right\}^2 \tag{12}$$

Formula (12), w_{im} it represents the intermediate layer information in backpropagation. The intermediate layer information in backpropagation is shown in formula (13).

$$w_{im} = -\eta \frac{\partial E}{\partial w_{mn}} \tag{13}$$

Formula (13), η represents the learning efficiency. After adjusting the weights, the root mean square error is further expanded, as shown in formula (14).

$$E_z = \frac{1}{2}\sum_{n=1}^{l}\left\{\left[y_n - f_2\left(\sum_{i=1}^{n} w_{im} f_{mn}\left(\text{int}_m\right)\right)\right]\right\}^2 \tag{14}$$

However, when only BPNN is used for calculation, there are problems with small local poles and insufficient convergence speed. The random search characteristic of the genetic algorithm is used to improve the convergence speed of BPNN. The basic flow of the genetic algorithm is shown in Figure 4.

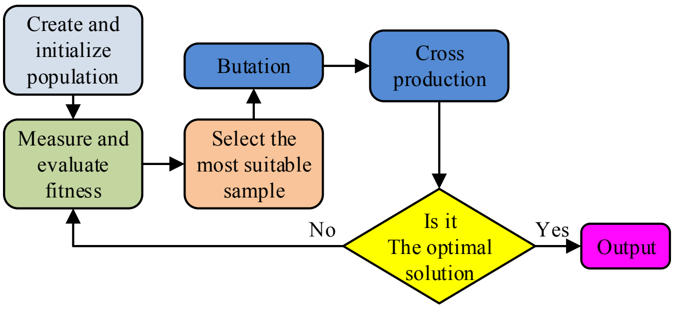

Fig. 4: Basic process of genetic algorithm

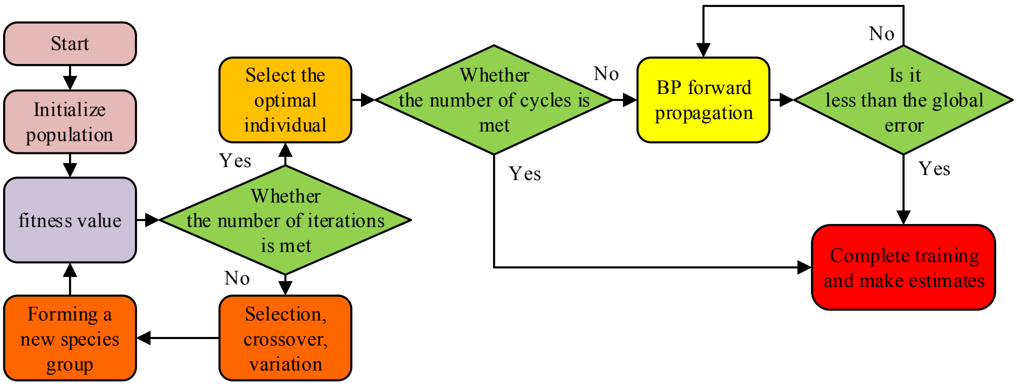

Fig. 5: Optimization of BPNN validation steps

It can be seen from Figure 4 that when the basic process of the genetic algorithm is running, it first needs to create and initialize the population and encode the parameter characteristics. Then measure and evaluate the adaptability, select suitable samples from the evaluation results, and judge whether the obtained results are the optimal solution after mutation and crossover operations. If the optimal solution is not reached, the adaptive measurement and evaluation are performed again, and then the loop operation is performed until the obtained result reaches the optimal solution, and the result is output. After combining the genetic algorithm with the BPNN, use the genetic algorithm to optimize the distribution of the weight thresholds of multiple targets. When encoding, use the real number encoding method to encode each individual. First, the input layer-hidden layer connection weight. Then encode the hidden layer-output layer connection weights, encode the neuron threshold of the hidden layer, and finally encode the neuron threshold of the output layer. When the population is initialized, the initial population number is set, and the initial value of the weight and threshold is defined as a real number between -1 and 1. The purpose of training is to make the cost estimate fit the actual value, and the absolute value of the error sum of the expected result and the predicted result is used as the optimization goal. After genetic manipulation, cost estimation is performed using optimal weights and thresholds. In the hidden layer of the neural network, the calculation of the number of nodes included is shown in the formula (15).

$$N_E = \sqrt{N_{E1} + N_{E0}} + t \tag{15}$$

Formula (15), N represents the number of nodes; t it is an integer between 1 and 10. The steps of verification according to the optimized BPNN are shown in Figure 5. It can be seen from Figure 5 that the optimization BPNN verification step starts with initializing the population. After generating the fitness value, if the preset number of iterations is not reached, the selection, crossover, and mutation operations are performed to form a new population, and then repeated iterations until after the number of iterations reaches the preset number of times, the optimal individual is selected. If the preset number of cycles is met, the training is completed. If not, the forward propagation is performed until the error is smaller than the global error.

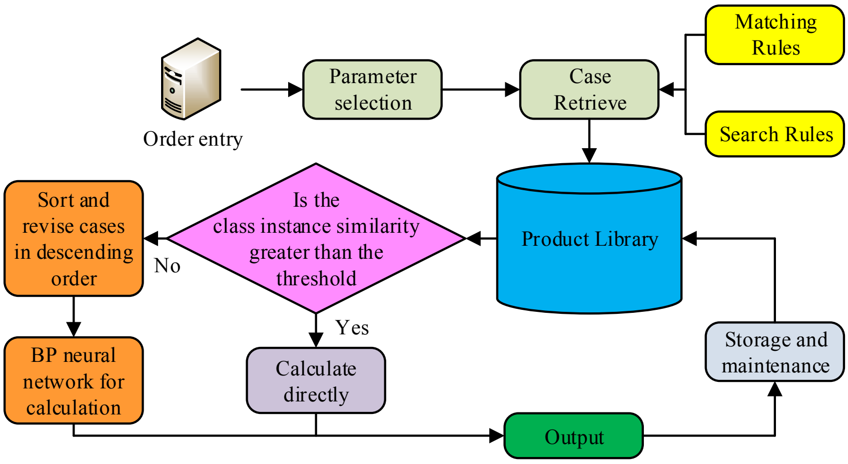

Fig. 6: Cost estimation methods for manufacturing enterprises

Integrating the extension matter-element technology into the cost assessment, the products involved in the assessment can be searched and matched quickly and accurately. Finally, the big data cost estimation method for manufacturing enterprises constructed is shown in Figure 6.

As can be seen from Figure 6, when performing cost estimation, firstly, the parameter selection of the model is performed based on the input order information, and the threshold value of the resulting unit is set, and then the instance retrieval is performed, and the representative class instances are corrected and then learned. If the similarity of the retrieved class instances is less than the threshold, the cases are sorted and corrected in descending order, and then calculated by the BPNN; if it is greater than the threshold, the calculation is performed directly. The calculation results are output and stored in the product library to expand the richness of search samples, and the output results are the cost estimation results.

4 Effectiveness Analysis of Manufacturing Enterprise Cost Estimation Method based on BPNN and Big Data Analysis

Manufacturing enterprise cost estimation can bring data reference for enterprise decision-making. This section will conduct performance tests and application analyses of the research method to determine the effectiveness of the research method.

4.1 Performance Test of Cost Estimation Method based on BPNN and Big Data Analysis

To analyze the effectiveness of the cost estimation method based on BPNN and big data analysis designed by the research in estimating manufacturing enterprises, the research first tests the performance of the designed method. The data set used in the test is a composite data set formed by mixing the historical cost data set and the external data set, and the data set is divided into two sub-blocks. The decision tree random forest algorithm is a method that allows for online model updates. The decision tree model provides a clear decision path, allowing users to understand how the model makes predictions; The random forest algorithm can perform data analysis without the need for a large amount of preprocessing and is a high-performance enterprise data analysis method. The support vector machine deep neural network algorithm has strong generalization ability and can adapt to the complex data distribution of manufacturing enterprises; The support vector machine algorithm can provide certain model interpretability, help users understand content, and has good performance in manufacturing-related data analysis. In this context, the study compared the decision tree random forest algorithm with the support vector machine deep neural network algorithm. Test the loss value during the training of the research method, as shown in Figure 7.

It can be seen from Figure 7 that during training, the loss values of the three different methods all gradually decrease in the early stage, and tend to stabilize after reaching the lowest interval.

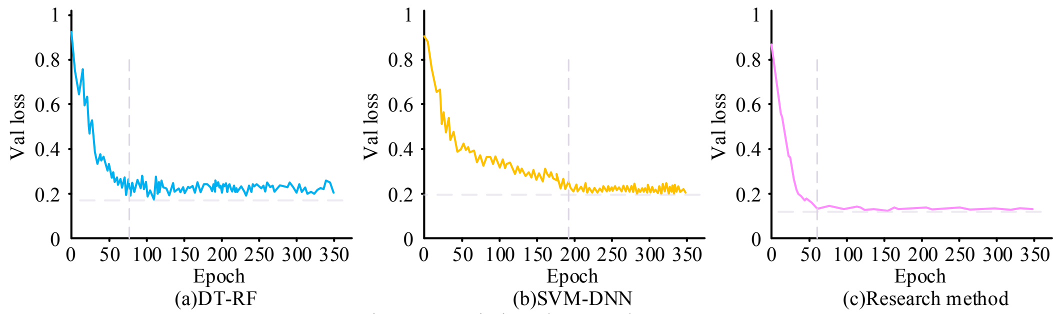

Fig. 7: Training loss value test

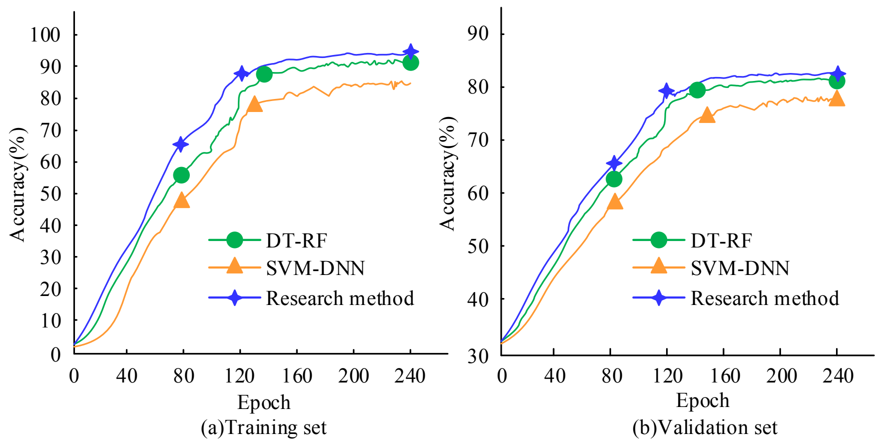

Fig. 8: Estimating accuracy testing

The loss value of the decision tree-random forest algorithm reaches the lowest interval after 77 iterations, and the curve has obvious fluctuations in the process of descending, and the lowest value of the interval is 0.17. The loss value of the support vector machine-deep neural network algorithm reaches the lowest interval after 191 iterations, and the curve has obvious fluctuations during the decline process, and the lowest value of the interval is 0.19.

The loss value of the research method reaches the lowest interval after 61 iterations, and there is a very small fluctuation in the lowest interval, and the lowest value of the interval is 0.12. It shows that the research method has faster training speed and better training results. The estimation accuracy of the research method was tested, as shown in Figure 8.

It can be seen from Figure 8 that in both the training set and the verification set, the estimation accuracy of the three methods increases with the number of iterations, and tends to be stable after reaching the highest interval. In the training set, the estimation accuracy of the decision tree-random forest algorithm increased rapidly during the first 138 iterations, and the estimation accuracy was 90.8% when the number of iterations reached 240. The estimation accuracy of the support vector machine-deep neural network algorithm increased rapidly during the first 129 iterations, and the estimation accuracy was 84.1% when the iteration number reached 240. The estimation accuracy of the research method increased rapidly during the first 120 iterations, and the estimation accuracy was 94.7% when the number of iterations reached 240. In the verification set, the estimation accuracy of the decision tree-random forest algorithm increases rapidly during the first 141 iterations, and the estimation accuracy is 81.7% when the number of iterations reaches 240. The estimation accuracy of the support vector machine-deep neural network algorithm increases rapidly in the first 148 iterations, and the estimation accuracy is 77.6% when the iteration number reaches 240. The estimation accuracy of the research method increased rapidly during the first 120 iterations, and the estimation accuracy was 83.2% when the iteration number reached 240. It shows that the research method has better estimation accuracy. The calculation time of the research method in different data scales is tested, as shown in Figure 9.

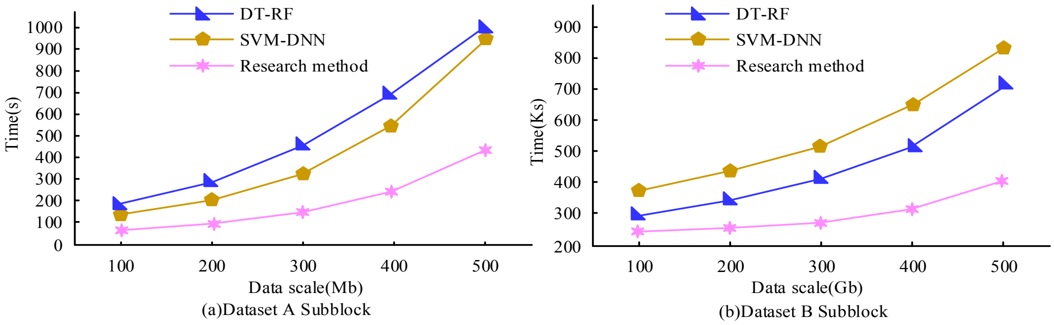

Fig. 9: Calculation time test

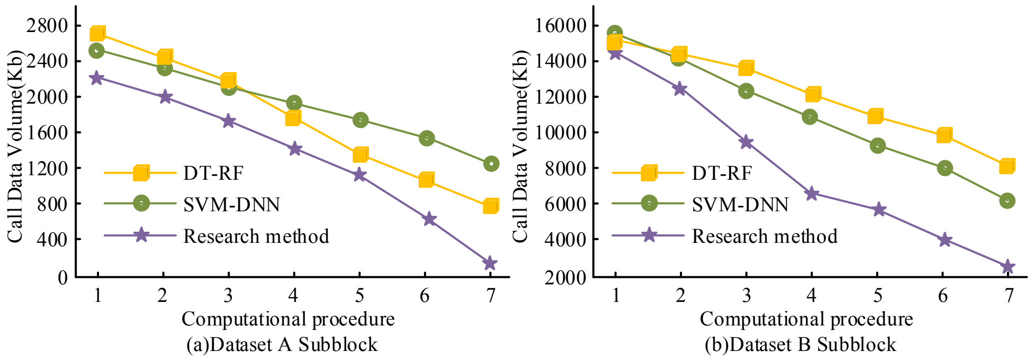

Fig. 10: Call data volume

It can be seen from Figure 9 that the calculation time of the three methods increases with the increase of the data size. In the A sub-block with a smaller data size, the decision tree random forest algorithm has a computation time of 181 seconds when the data size is 100 Mb and 998 seconds when the data size increases to 500 Mb. The calculation time of the support vector machine-deep neural network algorithm is 134 s when the data size is 100 Mb, and the calculation time is 946 s when the data size increases to 500 Mb. The calculation time of the research method is 69 s when the data size is 100 Mb, and the calculation time is 431 s when the data size increases to 500 Mb. In the B sub-block with a large data size, the calculation time of the decision tree-random forest algorithm is 289 Ks when the data size is 100 Gb, and the calculation time is 721 Ks when the data size increases to 500 Gb. The calculation time of the support vector machine deep neural network algorithm is 376s when the data size is 100Mb, and 830Ks when the data size increases to 500Gb. The calculation time of the research method is 244 Ks when the data scale is 100 Gb, and the calculation time is 403 Ks when the data scale increases to 500 Gb. It shows that the research method has better calculation speed. The recalled data volumes of the research method at different computational steps are tested, as shown in Figure 10. It can be seen from Figure 10 that the call data volumes of the three methods decrease continuously as the calculation proceeds. In the A sub-block with a smaller data size, the decision tree random forest algorithm has a data call volume of 2711Kb at step 1 and 762Kb at step 7. The data call volume of the support vector machine deep neural network algorithm in step 1 is 1522Kb, and the call data volume in step 7 is 1241Kb. The data call volume in step 1 of the research method is 2203Kb, and the call data volume in step 7 is 164Kb. In sub-block B with large data size, the calling data volume of the decision tree-random forest algorithm in the first step is 15213 Kb and the calling data volume in the seventh step is 8123 Kb. The call data volume of the support vector machine-deep neural network algorithm in the first step is 15653 Kb, and the call data volume in the seventh step is 6167 Kb. In the research method, the calling data volume in the first step is 1442 Kb, and the calling data volume in the seventh step is 2481 Kb. It shows that the research method has better data retrieval simplicity.

4.2 Application Analysis of Cost Estimation Method based on BPNN and Big Data Analysis

In the application analysis of the research method, an elevator parts manufacturing enterprise is taken as the analysis object. First, analyze the processor usage of the system when performing cost estimation, as shown in Figure 11.

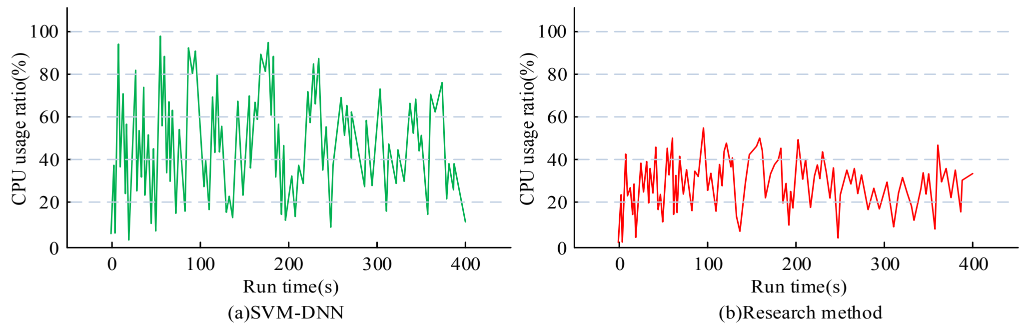

Fig. 11: Processor occupancy

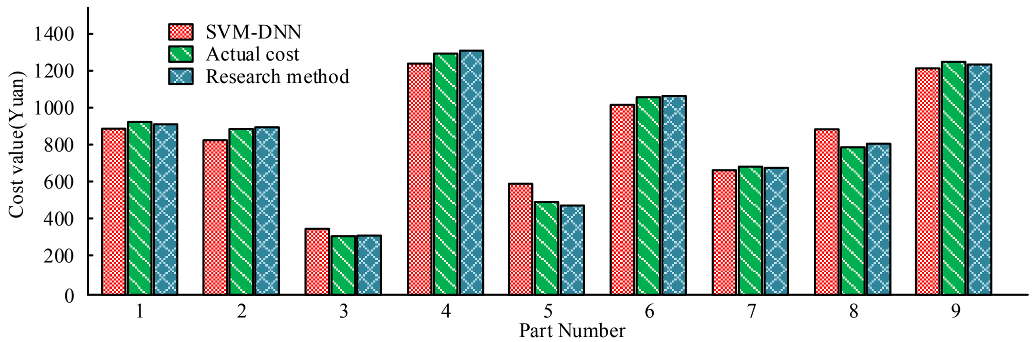

Fig. 12: Cost estimation results

From Figure 11, it can be seen that during cost estimation, there is a certain degree of fluctuation in the processor occupancy of different methods during a total runtime of 400s. The maximum processor utilization ratio of the support vector machine deep neural network algorithm reached 97%, and there have been multiple instances during this period where it approached the maximum processor utilization ratio; The average processor usage during this period reached 61%. The minimum processor occupancy ratio is the case of algorithm pauses or system protection, and does not have a reference value. The maximum processor utilization ratio of the research method is 55%, and the average processor utilization ratio during the period is 25%; Due to the relatively small maximum processor utilization ratio, the fluctuation in processor utilization ratio during the time period is also relatively small. It shows that the research method brings less burden to the processor during actual operation, and has lower requirements on hardware. The cost estimation results of the research methods are compared and analyzed, as shown in Figure 12.

It can be seen from Figure 12 that the cost of 9 target parts has been successfully estimated using the support vector machine-deep neural network algorithm and research method. The actual processing costs of these nine parts are all below 1,400 yuan. The minimum difference between the estimated results of the support vector machine-deep neural network algorithm and the actual processing costs is 22 yuan, and the maximum difference reaches 117 yuan; among these 9 target parts, there were 8 instances where there was a significant deviation from the actual processing cost. the minimum difference with the actual processing cost is 4 yuan, and the maximum difference is 19 yuan; among these 9 target parts, there was no significant difference between them and the actual processing cost. It shows that the research method can effectively and accurately estimate the cost of manufacturing enterprises. Sensitivity analysis was conducted on cost indicators using research methods, and seven common implicit costs were set as indicators. The results are shown in Table 1.

Table 1. Sensitivity Analysis of Implicit Cost Indicators

Index	Sensitivity value
Policy guarantee capability	-0.514
Market growth degree	-0.708
Ecological environment	-0.330
Supporting service industry development environment	-0.130
Government work efficiency	-0.218
Investment and financing environment	-0.526
Science and technology innovation ability	-0.617

From Table 1, it can be seen that the research method has obtained analysis results on the seven cost sensitivities of manufacturing enterprises. The absolute sensitivity value of the market development level is above 0.7, which is the highest among the 7 items, indicating that the cost of manufacturing enterprises is greatly influenced by market development level factors and has a strong sensitivity to changes in the market development level. The absolute sensitivity value of the development environment of the supporting service industry is 0.130, which is the smallest of the seven, indicating that the cost of manufacturing enterprises is less affected by changes in the development environment of the supporting service industry.

5 Conclusion

The cost estimation of manufacturing enterprises can provide a decision-making reference for the production planning of manufacturing enterprises. Based on big data analysis, the research proposes a cost estimation method for manufacturing enterprises using BPNN. Firstly, the construction of the big data analysis architecture is completed based on the Lambda architecture, and then the data association analysis is carried out through the clustering algorithm, and the cost is estimated using the optimal weight and threshold, and finally, the effectiveness of the research method is tested. The experimental results show that in the training loss value test, the loss value of the research method reaches the lowest interval after 61 iterations and the lowest reaches 0.12; in the estimation accuracy test, the research method in the verification set reaches 240 iterations. The estimation accuracy is 83.2%; in the calculation time test, the calculation time of the research method is 69 s when the data size in the small-scale data set is 100 Mb; in the analysis of processor occupation, the maximum processor occupation of the research method is 400 s. The ratio is 55%; the maximum difference between the cost estimation results of the nine target parts and the actual value by the research method is only 19 yuan. The results show that the research method has better computational efficiency and accuracy of results when estimating the cost of manufacturing enterprises, and the burden on the hardware is smaller. In the future, research methods can be applied to manufacturing enterprises with intelligent data collection equipment. The data collection equipment monitors and collects data on the production line, calculates and analyzes the collected production data through research methods, and obtains corresponding analysis results. Enterprise management personnel refer to the analysis results to adjust the production plan of the enterprise and make decisions on the development direction of the enterprise. However, research methods are more focused on designing for mechanical manufacturing enterprises, and data from mechanical manufacturing enterprises is also used for application analysis. Currently, it is uncertain how effective the application will be in areas where there are few intelligent data collection devices in the light textile industry and handicrafts industry. In the future, application analysis will be conducted for manufacturing categories with relatively small amounts of intelligent data collection to enrich experimental results and optimize research methods.

References:

[1] L. Mauler, F. Duffner, W. G. Zeier, J. Leker. "Battery cost forecasting: a review of methods and results with an outlook to 2050," *Energy Environ. Sci*., vol. 14, no. 9, pp. 4712-4739, 2021.

[2] C. Diagne, B. Leroy, A. C. Vaissière, R. E. Gozlan, D. Roiz, I. Jarić, J. Salles, C. Bradshaw, F. Courchamp. "High and rising economic costs of biological invasions worldwide," *Nature,* vol. 592, no. 7855, pp. 571-576, 2021.

[3] X. R. Zhang, X. Sun, W. Sun, T. Xu, P. P. Wang, S. K. Jha. "Deformation expression of soft tissue based on BP neural network," *Int. J. Autom. Soft Comput.*, vol. 32, no. 2, pp. 1041-1053, 2022.

[4] K. Schwarze, J. Buchanan, J. M. Fermont, H. Dreau, M. W. Tilley, J. M. Taylor, P. Antoniou, S. Knight, C. Camps, M. Pentony, E. Kvikstad, S. Harris, N. Popitsch, A. Pagnamenta, A. Schuh, J. Taylor, S. Wordsworth. "The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom," *Genet. Med.*, vol. 22, no. 1, pp. 85-94, 2020.

[5] B. Khan, W. Khan, M. Arshad, N. Jan. "Software cost estimation: algorithmic and non-algorithmic approaches," *Int. J. Data Sci. Adv. Analytics*, vol. 2, no. 2, pp. 1-5, 2020.

[6] F. Elghaish, S. Abrishami, M. R. Hosseini, S. Abu-Samra. "Revolutionising cost structure for integrated project delivery: a BIM-based solution," *Eng. Constr. Archit. Manage.*, vol. 28, no. 4, pp. 1214-1240, 2021.

[7] S. Mishra, M. N. Sahoo, A. Kumar Sangaiah, S. Bakshi. "Nature-inspired cost optimisation for enterprise cloud systems using joint allocation of resources," *Enterprise Inf. Syst.*,

vol. 15, no. 2, pp. 174-196, 2021.
[8] A. Fazeli, M. S. Dashti, F. Jalaei, M. Khanzadi. "An integrated BIM-based approach for cost estimation in construction projects," *Eng. Constr. Archit. Manage.*, vol. 28, no. 9, pp. 2828-2854, 2021.
[9] I. Nevliudov, I. Razumov-Fryziuk, V. Yevsieiev, D. Nikitin, D. Blyzniuk, R. Strelets. "Cost estimation of photopolymer resin for 3D exposure of circuit boards," *Technol. Audit Prod. Reserves*, vol. 2, no. 2 (64), pp. 43-49, 2022.
[10] P. Karar, M. Chatterjee, S. Deshi, M. Das, P. Das, S. Pal. "Cost estimation and fabrication of automatic hand sanitizing machine," *Int. J. Res. Eng. Sci. Manage.*, vol. 4, no. 5, pp. 24-27, 2021.
[11] J. Leelathanapipat, P. Paichit. "Cost estimate of repairing refurbished equipment using multiple regression model," *Eng. J. Res. Dev.*, vol. 31, no. 2, pp. 127-138, 2020.
[12] W. Rosa, B. K. Clark, R. Madachy, B. W. Boehm. "Empirical effort and schedule estimation models for agile processes in the US DoD," *IEEE Trans. Software Eng.*, vol. 48, no. 8, pp. 3117-3130, 2021.
[13] S. Song, X. Xiong, X. Wu, Z. Xue. "Modeling the SOFC by BP neural network algorithm," *Int. J. Hydrogen Energy*, vol. 46, no. 38, pp. 20065-20077, 2021.
[14] X. Li, J. Wang, C. Yang. "Risk prediction in financial management of listed companies based on optimized BP neural network under digital economy," *Neural Comput. Appl.*, vol. 35, no. 3, pp. 2045-2058, 2023.
[15] J. Liu, J. Huang, R. Sun, H. Yu, R. Xiao. "Data fusion for multi-source sensors using GA-PSO-BP neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 10, pp. 6583-6598, 2020.
[16] P. Nanglia, S. Kumar, A. N. Mahajan, P. Singh, D. Rathee. "A hybrid algorithm for lung cancer classification using SVM and Neural Networks," *ICT Express*, vol. 7, no. 3, pp. 335-341, 2021.
[17] S. Ma, C. Zhou, C. Chi, Y. Liu, G. Yang. "Estimating physical composition of municipal solid waste in China by applying artificial neural network method," *Environ. Sci. Technol.*, vol. 54, no. 15, pp. 9609-9617, 2020.
[18] L. Chen, V. Jagota, A. Kumar. "Retracted article: research on optimization of scientific research performance management based on BP neural network," *Int. J. Syst. Assur. Eng. Manage.*, vol. 14, no. 1, pp. 489-489, 2023.
[19] X. Rong, Y. Liu, P. Chen, X. Lv, C. Shen, B. Yao. "Prediction of creep of recycled aggregate concrete using back- propagation neural network and support vector machine," *Struct. Concr.*, vol. 24, no. 2, pp. 2229-2244, 2023.
[20] S. Nimrah, S. Saifullah. "Context-free word importance scores for attacking neural networks," *J. Comput. Cognit. Eng.*, vol. 1, no. 4, pp. 187-192, 2022.
[21] X. B. Zhang, G. D. Cheng, J. L. Zhao, W. Dai, Y. Tao. "Research on classification and evaluation of Chang 3 reservoir in H area based on BP neural network technology," *Prog. Geophys.*, vol. 38, no. 3, pp. 1272-1281, 2023.
[22] Z. Cui, L. Wang, Q. Li, K. Wang. "A comprehensive review on the state of charge estimation for lithium-ion battery based on neural network," *Int. J. Energy Res.*, vol. 46, no. 5, pp. 5423-5440, 2022.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Huijian Ma is the solely author who conducted the methodology, writing and revision of this manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflict of interest to declare.