

Research on Construction Project Cost Prediction Model based on Multilayer Perceptual Machine

Honghui Wang^{1, a}, Zihan Li^{1, b, *}, Ziyang Li^{2, c}

¹ School of Architecture and Construction, Lanzhou jiaotong University, Lanzhou 730000, China

² School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, China

^a 523581837@qq.com, ^{b, *} 1019741068@qq.com, ^c 913148074@qq.com

* Corresponding Author: Zihan Li

Abstract. Since China's construction project pricing has long been in the stage of quota and bill of quantities concurrently, most of the projects are still based on the quota in the pre-project cost measurement, which consumes a lot of human and material resources and financial resources. In this paper, we use the method of machine learning to firstly select the cost elements, use the method of t-SNE to do the dimensionality reduction of the data, and then predict the unilateral cost of the construction project according to the cost elements by establishing the multilayer perceptual machine model based on Adam optimization, and finally evaluate the performance of the model. The experimental results show that the accuracy of the model meets the needs of project pre-investment after the trained model.

Keywords: t-SNE Dimensionality Reduction; Engineering Cost; Cost Elements; Multilayer Perceptron.

1. Introduction

According to China's Bureau of Statistics data show that China's construction industry in 2022 added value of 8338.3 billion yuan, an increase of 5.5% over the previous year, the annual investment in real estate development of 132,895 billion yuan, the country has a qualification level of general contracting and professional contracting construction enterprise profit of 836.9 billion yuan, for the number of funds such a large number of investments, the project's pre-investment decision-making is particularly important, but through the research, most of the Owners in the project information is more perfect bidding stage is still using the quota, and even bidding quotes are mostly empirical values, in the project if you can be more accurate and fast to understand the cost of the project, can better help owners to make investment decisions. But when making investment decisions in the face of a small amount of project information, how can we quickly and accurately understand the cost of the project? In the rapid development of information technology, we can make full use of big data, artificial intelligence, machine learning and other methods to establish project cost prediction model, in the construction project can be quickly and efficiently predicted the cost of construction projects.

2. Literature Review

In terms of cost prediction, many scholars have conducted research using different model predictions. Jinhao Xie [1], used BP neural network model and BP model based on genetic algorithm (GA) to do comparative experiments, and found that the maximum prediction error of BP neural network is around 10%, while the prediction error of GA-BP model is within 5%. Dong Na et al.[2], by combining the artificial bee colony algorithm (ABC) and the support vector regression machine (SVM) to construct a prediction model for construction projects in the pre-project stage, the regression fit of the model $R^2 = 0.999$, and compare the ABC-SVM model with the GRID-SVM and BP neural network which comes with parameter optimization, the comparison results show that the root mean square error and the average relative error of the ABC-SVM model are both the same as

those of the GA-BP model. error and mean relative error are higher than the remaining two models. Prasetyono Puguh Novi et al.[3] argued that the fluctuation of construction price depends on the construction cost such as building materials and labor, so linear regression was used to predict the construction cost in Indonesia. Wang Yali et al.[4] proposed a new method by using regression analysis (RA) as a front-end prediction, and artificial neural networks (ANN) as a back-end correction, they proposed a RA-ANN cost prediction model based on historical data, which improves the accuracy of cost prediction in the pre-project stage where the amount of information is limited. The MLP model is also widely used in various fields. In earthquake prediction, Cuicui Xu et al. [5], applied the MLP approach to the evaluation of instrumental seismic intensity. The authors applied Relu activation function between four hidden layers and the hidden layer and the output layer. The authors applied Relu activation function between four hidden layers and Sigmoid activation function between hidden layer and output layer, and evaluated the model by MAPE after 1000 iterations, which proved the superiority of the MLP model through example verification and comparison with other methods. In the field of aviation, Wang Xumin et al.[6] , applied the MLP model in the prediction of stress temperature of aero-engine compressor disk, and when Relu was selected as the activation function, and the proposed Newton method (LBFGS) was used as the optimization strategy, the relative deviation between the training set and the validation set was less than 1%, and a comparison was made with the traditional finite element computation, and it was found that the MLP model could improve the computational efficiency while guaranteeing the high accuracy. . In the field of agriculture, Chen Yifei et al. [7] introduced the MLP model with 3-layer neural network for dynamic nonlinear regression prediction in the study of the variation of tomato diameter diameter in room temperature, and the experimental results showed that the regression coefficient (R2) of the model was 0.901, and the root mean square error (RMSE) was 0.175, which is a good scenario for application. In the field of medicine, Zeinab Kohzadi et al[8],used multilayer perceptron to determine the severity of sleep apnea as a disease, designed a multilayer perceptron containing two hidden layers, performed 120 iterations by Sigmoid activation function and trained the model using 10-fold cross-validation, and the experimental results showed that the classification correctness of the multilayer perceptron was 91% The results show that the multilayer perceptron can classify correctly at 91% or more.

This paper applies multilayer perceptron to the study of the marketization of construction project cost elements, and migrates the regression problem in data mining to the prediction of construction project cost elements so that owners can quickly and accurately know the marketed unilateral cost when making investment decisions.

3. Selection of Cost Elements

3.1. Initial Selection of Cost Elements

Table 1. Preliminary Selection of Project Cost Characteristics

serial number	Characteristic Primary Collection	serial number	Characteristic Primary Collection
1	Floor area	9	Doors and windows
2	Underground floor area	10	HRB400 Composite Price
3	Number of floors above ground	11	Average market labor costs
4	basement layer	12	Average market cost of materials
5	Average floor height	13	Average market machinery costs
6	eaves height	14	time allocated for a project
7	Structure type	15	GDP
8	base type	16	Engineering Cost Index

This paper combined with the Ministry of Housing and Construction issued the "project cost index classification and compilation guide" in the house building construction project characteristics of information and in recent years, some scholars [9-13] on the construction project cost prediction of the features selected, now the preliminary selection of features are summarized and collated in the table 1.

The cost elements selected in this paper do not include construction period interest, fees, taxes, and other cost factors that have little correlation with changes in market information.

3.2. Correlation Coefficient Analysis

The correlation coefficient is to transform the variables into standard units, and it mainly solves the problem of different scale between the variables, but in the actual engineering projects, the same cost elements may have different units, so it is necessary to unify the scale when doing machine learning. Since there are not only continuous variables but also discrete variables in actual construction projects, this paper uses two methods, Pearson correlation and Spearman rank correlation coefficient, to analyze the correlation between the data. The correlation coefficient expression and the correlation degree are indicated in Figure 1 as follows:

$$\rho_{xy} = \frac{\text{Cov}(x, y)}{\sqrt{D(x)}\sqrt{D(y)}} \quad (1)$$

$C_{ov}(x, y)$: x , covariance of y ; $D(x)$: standard deviation of x ; $D(y)$: standard deviation of y .

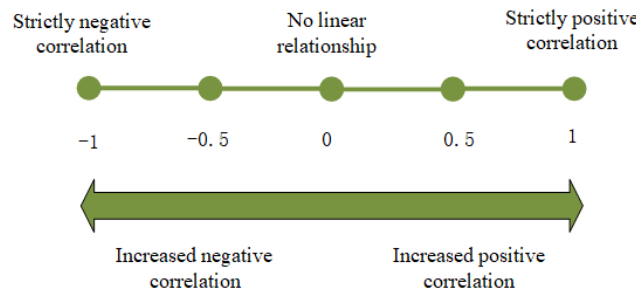


Fig 1. Graphical representation of the degree of correlation

3.2.1. Pearson Correlation Coefficient Analysis

Pearson correlation coefficient, also known as Pearson product-moment correlation coefficient (Referred to as PPMCC or PCCS) [14]. The Pearson correlation coefficient is used when the variables are from the same project, there is a linear relationship between the variables, and the variables themselves are normally or nearly normally distributed. The heat map of Pearson correlation coefficient is shown in Fig. 2 and its expression is as follows:

$$r = \frac{\sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{S_x S_y}}{n-1} \quad (2)$$

Where $x_i - \bar{x}$: the deviation of x , S_x : the standard deviation of x , $y_i - \bar{y}$: the standard deviation of y , S_y : the standard deviation of y .

3.2.2. Spearman's Rank Correlation Coefficient

Spearman's rank correlation coefficient, also commonly known as Spearman's rank correlation coefficient or rank correlation, "rank" can be interpreted as a kind of order or sorting. This method is generally used for the analysis of the correlation between the name data and order data and other non-continuous variables with the nature of the rank, which does not require high data distribution, in the face of the data does not obey the normal distribution or the distribution of the data of unknown type,

this is also a commonly used correlation analysis method. The heat map of Spearman's rank correlation coefficient is shown in Figure 3, and its expression is as follows:

$$r_s = \frac{\sum (R_x - \bar{R}_x)(R_y - \bar{R}_y)}{\sqrt{\sum (R_x - \bar{R}_x)^2 \sum (R_y - \bar{R}_y)^2}} \quad (3)$$

The simplified formula is

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad (4)$$

Where: R_x : the rank of x , $\bar{R}_x$: the average rank of x , R_y : the average rank of y , $\bar{R}_y$: the average rank of y , n : the number of samples, d_i : the difference in rank between the data.

The strength of correlation determination table for Spearman's rank correlation coefficient is shown in Table 2 below:

Table 2. Correspondence between Spearman's rank correlation coefficients and degree of correlation

Spearman's rank correlation coefficient	degree of relevance	Spearman's rank correlation coefficient	degree of relevance
0.0-0.2	weak	0.6-0.8	vigorous
0.2-0.4	slightly less than	0.8-1.0	eager get ahead in life
0.4-0.6	moderate		

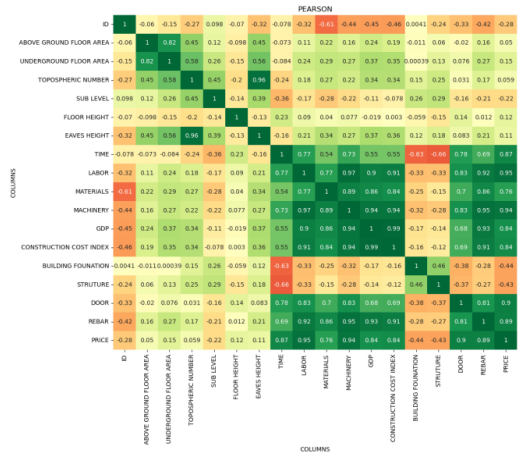


Fig 2. Heat map of Pearson's correlation coefficient

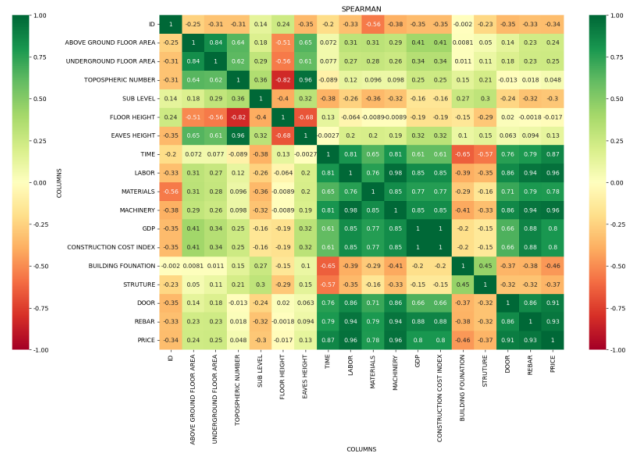


Fig 3. Heat map of Spearman's rank correlation coefficient

The analysis of Pearson and Spearman rank correlation coefficients revealed that the correlation between duration, average market labor cost, average market material cost, average market machinery cost, structural type, foundation type, GDP, engineering cost index, door and window volume, and HRB400 composite price were found to be stronger, and they were selected as the cost elements for predicting the project cost.

3.3. Establishment of Cost Element System

The analysis of the Pearson and Spearman rank correlation coefficients resulted in the classification of cost elements into two categories, technology-based and market-based elements, as shown in Table 3 below.

Table 3. Final selection of engineering cost elements

Typology	Element name	Thrust
Technique Technology Type Essentials Element	Structure type	Types of building load-bearing or maintenance structures
	Base type	An enlarged structure of a building buried below ground level to carry the loads transmitted by the building above it.
	Doors and windows	Door and window area as a percentage of wall area
Market Type Elements	Average market labor costs	Average value of labor costs at market prices over the duration of the construction project
	Average market cost of materials	Average cost of materials at market prices over the duration of the construction project
	Average market machinery costs	Average value of machinery costs at market prices over the duration of the construction project
	HRB400 Composite Price	Comprehensive price of HRB400 rebar
	time allocated for a project	Time taken from official groundbreaking to commissioning of the project
	GDP	GDP value of the year in which the project was carried out
	Engineering Cost Index	Ratio of construction project cost reporting period to base period

4. Modeling

4.1. Data Pre-processing

Because in the actual engineering projects, the data difference between different projects is large, that is, the scale of the project is relatively large, part of the data will also be very large, while the scale of the project is small, part of the data will also be small, is now analyzed on its raw data (as an example of unilateral cost): from the histogram and continuous probability density estimation in Fig. 4, we can see that the distribution of unilateral cost does not obey the normal distribution, the existence of the phenomenon of bimodal peaks, and the There is the problem of heteroskedasticity, analyze the reason may be affected by the epidemic, resulting in human materials and machinery and other costs rise faster so that the gradual growth of the price of the phenomenon of discontinuous growth; from the box plot in Figure 4 can be seen that the box distribution is skewed to the left, and the box is pressed flat, indicating that the distribution of the target variable is not uniform, and the same may be the existence of anomalies. Therefore, the data were log-transformed to alleviate the problem of heteroskedasticity and reduce the impact of outliers on the sample as a whole. After the logarithmic transformation, although the histogram and the continuous probability density estimation plot in Figure 5 also have the phenomenon of bimodal peaks, the size of the bimodal peaks is significantly different, and the distribution between the data approximates a normal distribution, and the boxes of the box plots in Figure 5 are also roughly centered, mitigating the interference of the outliers as well as the differences in the magnitudes between the data. The raw data are shown in Figure 4, and the processed data are shown in Figure 5.

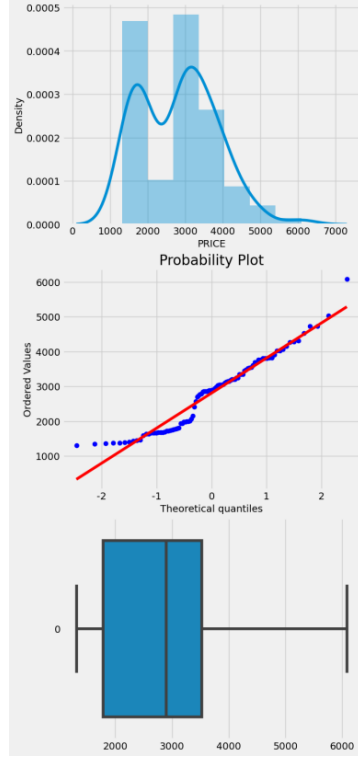


Fig 4. Histogram of raw data, continuous Probability density estimates and box plots

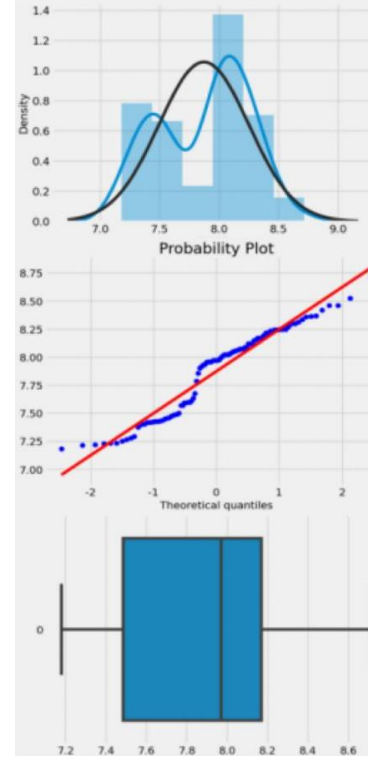


Fig 5. Histogram after logarithmic treatment, continuous Probability density estimates and box plots

4.2. t-SNE Dimensionality Reduction

There are many factors affecting the cost of a single party in the actual project, if all the factors are listed for analysis not only the amount of calculation is large, which will produce high computational costs and consume resources, but also produce a lot of redundant data, which will have a certain impact on the results of the experiment, and even produce a dimensional explosion, which will lead to the final model cannot be converged. Therefore, we have to choose the cost elements that have a greater impact on the target variables and do not generate redundant data for analysis, and at this time we have to use the method of dimensionality reduction.

t-SNE dimensionality reduction, i.e., t-distributed stochastic neighbor embedding (t-distributed stochastic neighbor embedding), is a dimensionality reduction method that transforms high-dimensional data into low-dimensional data in a nonlinear way [15]. Basic principle: Each data point is mapped to the corresponding probability distribution by mapping transformation. Specifically, a Gaussian distribution is used to convert the distance to a probability distribution in the high-dimensional space, and a t-distribution is used to convert the distance to a probability distribution in the low-dimensional space, so that the intermediate and low distances in the high-dimensional space can have a larger distance after the mapping, which makes the dimensionality reduction able to avoid focusing on the local features too much and ignoring the global features.

Commonly used principal component analysis attempts to map all clusters as a whole when dimensionality is reduced because local structure may be lost. t-SNE attempts to preserve the local structure of the data by minimizing the Kullback-Leibler scatter (KL scatter) between the two distributions with respect to the location of the midpoint of the map.

The stochastic nearest neighbor embedding algorithm (i.e., the traditional SNE dimensionality reduction algorithm) can convert the Euclidean distance between high-dimensional data points into a conditional probability characterizing the similarity [16]. The main idea is as follows: first, given the

data points in the high-dimensional space: $x_i(x_1, x_2, \dots, x_n)$, calculate the conditional probability of selecting x_j as the nearest neighbor centered on x_i with Gaussian distribution.

$$p_{j|i} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_i - x_k\|^2 / 2\sigma_i^2)} \quad (5)$$

There is $y_i(y_1, y_2, \dots, y_n)$ in the low-dimensional space corresponding to $x_i(x_1, x_2, \dots, x_n)$ in the high-dimensional space, calculate the conditional probability of choosing y_j as the nearest neighbor centered on y_i with a Gaussian distribution.

$$q_{j|i} = \frac{\exp(-\|y_i - y_j\|^2)}{\sum_{k \neq i} \exp(-\|y_i - y_k\|^2)} \quad (6)$$

In order to be able to keep the points in the higher dimensional space mapped to the lower dimensional space as much as possible the same distribution, i.e., the points that are far away in the higher dimensional space are also far away after mapping, and the points that are close to each other are also close to each other after mapping, the KL dispersion is used as a measure of this distance

$$C = \sum_i KL(P_i \| Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \quad (7)$$

At this point the problem is transformed into minimizing the above loss function C

$$\frac{\partial C}{\partial y_i} = 2 \sum_j (p_{j|i} - q_{j|i} + p_{i|j} - q_{i|j})(y_i - y_j) \quad (8)$$

However, KL distance has asymmetry, when $p_{j|i}$ is smaller and $q_{j|i}$ is larger, the Gaussian value at this time is smaller. That is, when the points in the high dimensional space are far away, when the points mapped to the low dimensional space are close, the penalty is small at this time, which is the opposite of the result we want. Therefore, firstly, we need to change the asymmetric SNE to symmetric SNE, and secondly, we need to change the Gaussian distribution in the low-dimensional space to t-distribution, in order to reduce the distance between the same kind of data, and increase the distance between different kinds of data, so we introduce the t-distribution stochastic nearest neighbor embedding algorithm (t-SNE). The distribution of cost elements without dimension reduction is shown in Fig. 6, and the distribution of cost elements after dimension reduction is shown in Fig. 7.

To avoid the effect of outliers, define the joint distribution in the high-dimensional space and change the distribution function in the low-dimensional space to a t-distribution

$$p_{ij} = \frac{p_{i|j} + p_{j|i}}{2n} \quad (9)$$

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_i\|^2)^{-1}} \quad (10)$$

At this point the loss function under KL scatter is

$$C = KL(P \| Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (11)$$

The gradient is

$$\frac{\partial C}{\partial y_i} = 4 \sum_j (p_{ij} - q_{ij})(y_i - y_j)(1 + \|y_i - y_j\|^2)^{-1} \quad (12)$$

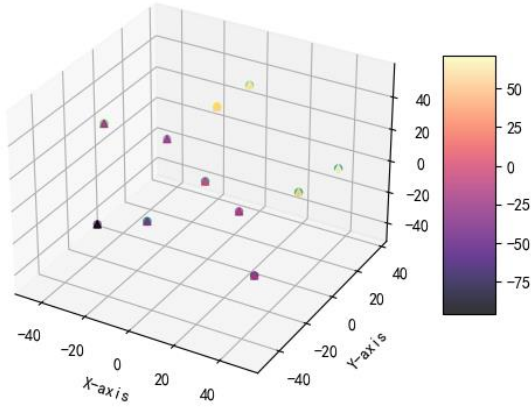


Fig 6. Three-dimensional spatial distribution of cost elements without downscaling

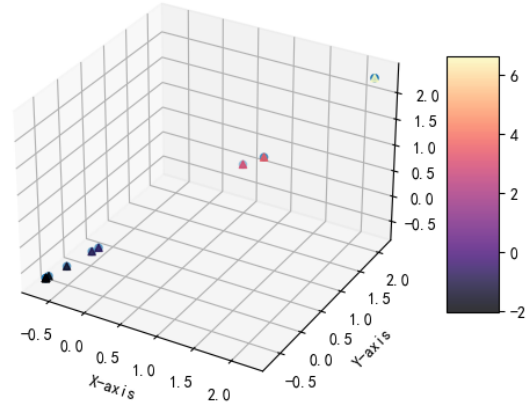


Fig 7. Three-dimensional spatial distribution of cost elements after dimensionality reduction

4.3. Establishment of a Multilayer Perceptual Machine

Multilayer Perceptron (MLP), a supervised feed-forward model in Artificial Neural Network (ANN), mainly consists of an input layer, a hidden layer, and an output layer. Output layer) [17,18]. Since the multilayer perceptron introduces a hidden layer compared to the original single-layer perceptron, it is also called Deep Neural Networks (DNN). The structure of the multilayer perceptron model is illustrated in Fig. 8.

The nodes of the input layer are the feature parameters processed in the previous section, and there are 10 of them. The input layer does not carry out any computation when it is connected with the hidden layer, and each of its nodes is fully connected with each node of the hidden layer according to different parameters. The hidden layer and the output layer are also connected in a fully connected manner. The simplest multicomponent perceptron has only one hidden layer, and assuming that the batch size of the sample X is n and the number of inputs is d , then the output of the hidden layer in this multilayer perceptron is:

$$H = XW_h + b_h \quad (13)$$

The output of the output layer is:

$$\begin{aligned} O &= HW_o + b_o = (XW_h + b_h)W_o + b_o \\ &= W_hW_oX + b_hW_o + b_o \end{aligned} \quad (14)$$

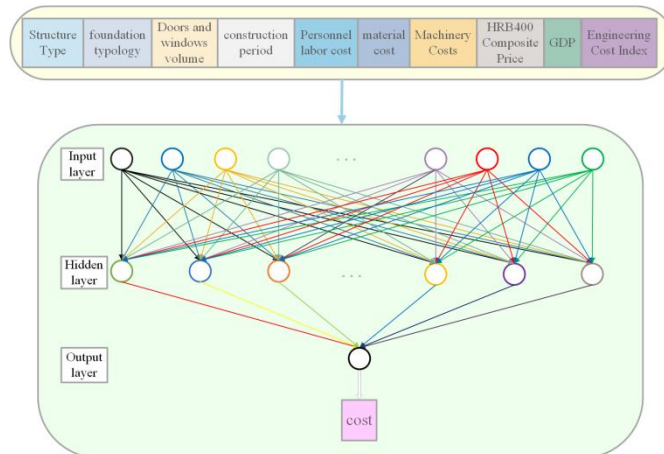


Fig 8. Structure of the multilayer perceptron model

4.3.1. Activation Function

Since the input layer and the hidden layer, the hidden layer and the output layer are all fully connected, and the fully connected layer is only an affine transformation of the data, multiple affine transformations superimposed on each other is still an affine transformation, that is, the data is still a linear relationship, but in real engineering, the relationship between the parameters is not completely linear, and therefore the introduction of nonlinear transformations --Activation function.

Commonly used activation functions are Sigmoid function, Tanh function (hyperbolic tangent function) and Relu function. Sigmoid function of the input mapping of the interval $(0, 1)$ [19], but its convergence is slow, when the input is very large or very small and its derivative is close to zero, which can easily lead to the disappearance of the gradient; the output of the function is not centered on 0, the weights of the update is slow; the output is close to linear when the input is near zero. When the input is close to zero, its output is close to linear transformation. Tanh function maps the input to the interval $(-1, 1)$, although its function image is similar to that of the Sigmoid function, the mean value of its output is centered on 0, which makes up for the disadvantage of the Sigmoid function that the mean value of its output is 0.5, when the input is large or small, the same problem of output smoothing and smaller gradient exists. Relu function is a segmented function that maps the input to the $(0, \infty)$ interval, its computation speed is faster than the Sigmoid function and Tanh function, and make up for the former gradient is smaller or the gradient disappears the problem, so the Relu function is the most used function. The images of the Sigmoid function, the Tanh function, and the Relu function are shown in Figs. 9, 10, and 11.

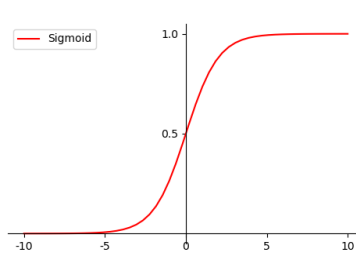


Fig 9. Sigmoid function

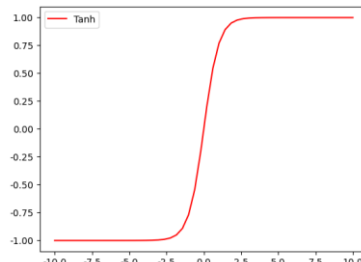


Fig 10. Tanh function

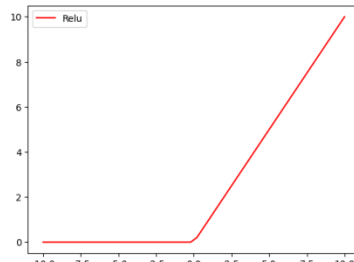


Fig 11. Relu function

In this paper, the Relu function is chosen as the activation function of the hidden layer to process the data. Theoretically the more layers of the hidden layer, the better the model fitting ability, but in practice it may lead to the problem of overfitting the model, and also increase the difficulty of training the model, so that the model is difficult to converge, and the more the number of nodes, the model learns to the depth of the shallow, and cannot be called deep learning.

4.3.2. k-fold Cross Validation

Data mining techniques require a large amount of data, and the data is divided into two independent sets when model training is performed: the training set and the test set, the training set is a collection of data to train the model, and the parameters of the model are adjusted through the training set to make the model have higher accuracy, find better hyperparameters, weights, etc. The trained model is applied to the test set to evaluate the performance of the model, in this paper, 20% of all the data is randomly selected as the test set and the remaining 80% of the data is used as the training set.

Due to the small amount of data, we have at present, when using a small number of samples to train the model there may be an overfitting phenomenon where the model overfits the data in the training set and the model's accuracy is lower on other data sets, the k-fold cross-validation method can be applied to tune the model's parameters by reusing the data independently and find the hyperparameters that make the model's generalization optimal. When the stability of the model is low or the data set is relatively small, a larger k-value is generally chosen and the model is trained several times, which can get a smaller deviation but will increase the time cost; when the stability of the

model is better or the training set is relatively large, the k-value is reduced to reduce the time cost of the model to be repeatedly fitted on different data blocks.

Figure 12 is an example of 10-fold cross-validation, where the data are evenly divided into 10 parts, and only 1 part is used as the test set for each training, and the remaining 9 parts are used as the training set, which is trained independently for 10 times, until each part of the data is made into a test set for 1 time, and the training set for 9 times is finished.



Fig 12. 10-fold cross validation

Each sample will get a predicted value after the training of the model, the error between the predicted value and the real value is called loss, the loss function is the function used to represent the loss [20]. The mean squared error loss function (MSE) is the function that represents the mean of the sum of the squares of the differences between the actual and predicted values. It is located between forward propagation and backward propagation [21], the so-called forward propagation is the process of inputting features into the model to get the predicted value, because there is an error between the predicted value and the real value, we hope that this error is as small as possible, so we need to back-propagate the data to update the parameters of the model, and this process is called back-propagation. Its role is to receive the predicted value of the model, calculate the error between the predicted value and the true value, and provide input data for backpropagation.

When making investment decisions, we hope that the cost prediction is highly accurate and fast, i.e., the smaller the value of the loss function, the smaller the error between the predicted value and the real value of the cost measurement, and the more accurate the model. At this time, it is necessary to use the Gradient Descent (Gradient Descent) algorithm to optimize the loss function, through the iterative way to find the minimum value of the loss function, or converge the loss function to the minimum value. In this paper, 10 cost elements are selected to train the model, some are positively or negatively correlated with the cost of single-placed, the loss function may not be a convex function, and the loss function using the commonly used Stochastic Gradient Descent (SGD) optimizer may fall into the local optimum rather than the global optimum, the optimizer used in this paper is the one firstly developed by Open AI's Diederik Kingma and Jimmie K. Kingma of the University of Toronto, and it is used to optimize the loss function by using the Gradient Descent algorithm. Diederik Kingma of Open AI and Jimmy Ba of the University of Toronto in a paper submitted to the 2015 ICLR (Adam: A Method for Stochastic Optimization) proposed a jack-of-all-trades optimizer: Adaptive Moment Estimation (AME), which combines the Adaptive Gradient Algorithm (Adagrad) and the Root Mean Square (RMS) algorithm. Adagrad) and Root Mean Square Propagation (RMSProp) together, by calculating the first-order moment estimation of the gradient to add inertia, so that the steeper the loss function, the faster the convergence speed, the second-order moment estimation proposes the adaptive learning rate, designing independent adaptive learning rates for different parameters, which can improve the performance on sparse problems, and can also be adjusted according to the average of the most recent weighting gradient, and performs well on-line and non-smooth problems, and can be said to be a good optimizer on sparse problems. smooth problems perform well and can be considered as a momentum version of PMSProp. As opposed to the traditional stochastic gradient descent method that keeps the same learning rate for each grid parameter and adjusts it individually as the learning

unfolds, Adam's algorithm sets different learning rates and different momentums for different parameters. Its formula is derived as follows:

Calculate the gradient of the loss function with respect to the parameters:

$$g_t = \nabla \hat{L}(\theta_t) \quad (15)$$

Calculate the first-order moment estimate of the gradient in momentum form (exponentially weighted moving average reflecting the gradient) at the moment:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \quad (16)$$

where m has an initial value of 0 and β_1 is the exponential decay rate of the first order moment estimate

Calculate the second-order moment estimate of the gradient in momentum form (exponentially weighted moving average reflecting the square of the gradient) at the moment:

$$\gamma_t = \beta_2 \gamma_{t-1} + (1 - \beta_2) g_t^2 \quad (17)$$

where γ has an initial value of 0 and β_2 is the exponential decay rate of the second-order moment estimate

Calculate the bias-corrected first-order moment estimates:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (18)$$

Calculate the bias-corrected second-order moment estimates:

$$\hat{\gamma}_t = \frac{\gamma_t}{1 - \beta_2^t} \quad (19)$$

Final update of parameters

$$\theta_{t+1} = \theta_t - \frac{n}{\sqrt{\hat{\gamma}_t} + \varepsilon} \hat{m}_t \quad (20)$$

where ε is a constant that tends to 0 to avoid dividing by 0 when updating parameters.

5. Model Acceptance and Comparison

The MLP model based on Adam optimization in this paper incorporates t-SNE dimensionality reduction, correlation coefficient analysis, and k-fold cross-validation to find the optimal activation function, the number of hidden layer layers, the number of neurons, the learning rate, and other optimal parameters after the selection of cost elements.

5.1. Model Acceptance

Since the learning rate, the number of training times and the number of folds for cross-validation need to be set in advance, for the learning rate, a larger value will make the model learn very fast initially, and a smaller value will make the model converge to a better performance; for the number of training times, the model may overfit if the number of training times is too high, and if the number of training times is too low, the model may be finished before it learns how to train; in general the folds for cross-validation are as long as they are smaller than the number of data in the number of data in the dataset can be less, although the more folds of cross-validation, the better the performance of the model, but it will waste a lot of resources[22]. The learning rate of the model, the number of cross-

validation folds and other parameters are analyzed through comparative tests, and finally the accuracy of the model is higher when the parameters are selected in the table below.

Table 4. MLP model parameter settings

Parameter name	parameterization	Parameter name	parameterization
activation function	Relu	Cross-validation folds	1000
Number of hidden layers	2	Number of iterations	10
Number of neurons in the hidden layer	21	Number of training sessions	1000

5.2. Model Comparison

In addition to the neural network model, a large number of scholars are exploring the accuracy of other prediction models in the face of the lack of engineering information at the pre-project stage, such as: Linear Regression [3], XGBoost [23], Ridge Regression [24], Support Vector Regression [25], etc. In this paper, we select two indicators to judge the accuracy of the model: the coefficient of determination R^2 (R-Square) and the root mean square error (RMSE Root Mean Squared Error). Square) and RMSE Root Mean Squared Error (RMSE) two indicators to judge the accuracy of the model, after experimental analysis found that: the MLP model is more accurate in the prediction of cost in the pre-project stage of construction engineering projects, and it is more suitable for the prediction of cost in this stage.

Table 5. Comparison of prediction errors of different models

Model name	R^2	RMSE
MLP	0.92	1437.45
Linear Regression	0.83	2153.40
XGBoost	0.87	1950.90
Ridge Regression	0.90	1639.15
Support Vector Regression	0.91	1528.73

6. Conclusion

In this paper, more than 100 engineering projects are selected for analysis, and 10 factors affecting the project cost are chosen, and they are used as inputs to the MLP model, and the unilateral cost is used as the output of the model, and a variety of methods are used to train and optimize the model, so that the model not only has an adaptive learning rate, but also a small number of samples can be used to make the model converge quickly, and the experiments have proved that the model has high precision, and the prediction accuracy Experiments have proved that the model has high accuracy and the prediction accuracy meets the requirements of investment decision-making, and it can adapt to the changes in the market price of manpower and machinery, providing a basis for the owner's investment decision-making.

References

- [1] Xie J. H, Liu W.C. Research on construction project cost prediction based on GA-BP neural network. Construction Economy, (2022)43(S1):235-240. DOI:10.14181/j.cnki.1002-851x. 2022 S10235.
- [2] DONG N, LU S.H, XIONG F. Construction project cost prediction based on ABC-SVM in the context of big data. Technical Economy, (2021)40(08):25-32.
- [3] Prasetyono, Pugh N; Mas S, H. S.; Dani, H. Predicting Construction Cost Using Regression Techniques for Residential Building. Journal of Physics: Conference Series, (2021) Vol.1899 No.1: 012120.
- [4] Wang, Y.L et al. Cost prediction of building projects using the novel hybrid RA-ANN model. Engineering, Construction and Architectural Management. (2023).

- [5] XU C.C, YANG S.J, XU W.X. Research on instrumental seismic intensity assessment method based on multilayer perceptron. *Journal of Disaster Prevention and Mitigation Engineering*. (2023) [https://doi.org/ 10.13409/ j.cnki. jdpme.20220823001](https://doi.org/10.13409/j.cnki.jdpme.20220823001).
- [6] Wang X.M Xu J.P, He Y, Stress-temperature prediction of aero-engine compressor disk based on multilayer perceptron. *Journal of Aerospace Dynamics*. (2022) <https://doi.org/10.13224/j.cnki.jasp.20220297>.
- [7] Chen Y.F. A prediction model of tomato stem diameter changes in greenhouse based on multilayer perceptron. *Xinjiang Agricultural Science*, (2020) 57 (3): 562-571.
- [8] Kohzadi Z, Safdari R, Haghighi K S. Determination of Sleep Apnea Severity Using Multi-Layer Perceptron Neural Network. *Sleep Medicine Research*, (2020) 11(2): 70-76.
- [9] Liang X, Liu Y. Construction project cost prediction model based on fuzzy neural network. *Technical Economy*, (2017) 36(03):109-113.
- [10] SHEN J.H, WANG S, SUN X.N. Research on cost prediction in the design phase of construction project based on LS-SVM. *Construction Technology*, (2018) 49(02):209-212.
- [11] Li W. Multivariate structural holistic linear regression model for construction cost prediction. *Construction Technology*, (2015) 46(09):846-849.
- [12] Park U; Kang Y.H; Lee H; Yun S. A Stacking Heterogeneous Ensemble Learning Method for the Prediction of Building Construction Project Costs. *Applied Sciences*, (2022) Vol.12(9729): 9729.
- [13] Shoar S; Chileshe N; David J.E. Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: application of random forest regression. *Journal of Building Engineering*, (2022) Vol. 50: 104102.
- [14] Shen J; Li L.L Lin D.M. Prediction of Engineering Parameters Based on Improved Artificial Neural Network. *IE&EM* (2019) 21 -27.
- [15] Wang S.T etc. A novel CC-tSNE-SVR model for rapid determination of diesel fuel quality by near infrared spectroscopy. *Infrared Physics & Technology*, (2020) Vol.106: 103276.
- [16] Meng F.L, Wang X.D, Wang Y.C et al. Research on horizontal displacement prediction of foundation pit based on TSNE dimensionality reduction algorithm. *Journal of Zhejiang University of Technology*, (2023) 51(04):403-411.
- [17] Tang Q.H, Li P.P, Li S.S. Research on cost prediction of residential projects based on multilayer perceptron. *Value Engineering*, (2019) 38(14):1-4.
- [18] Zheng Q.Z, Zhang Y, Xu S.S. Research on project cost prediction of LNG receiving station based on machine learning. *Construction Economy*, (2022) 43(S2):118-122.
- [19] Chen X. L. Research on cost prediction of high-rise residential project based on PSO-BP neural network. Lanzhou: Lanzhou Jiaotong University. (2020).
- [20] TIAN Y. Q, GUO G.Z. Estimation of cost of finishing works based on fuzzy prediction. *Science and Technology Innovation*, (2017) (25):82-83.
- [21] Guo K. Research on heating load prediction of residential buildings based on tree model and neural network model. Tianjin: Tianjin University, (2019).
- [22] Fan H. B. Research on construction project cost estimation model based on optimization neural network. Shenyang: Shenyang University of Architecture. (2019).
- [23] Guo J. J, Liu Y.W. Research on highway project estimate prediction method based on XGBoost. *Highway Transportation Science and Technology*, (2023),40(3): 58-68.
- [24] Liu Y, Lai X.X. Research on civil engineering investment estimation method based on multiple linear regression and engineering cost inversion[J]. *Journal of Qinghai Normal University-Natural Science Edition*, (2010) (1): 92-97.
- [25] Darko A etc. Using Machine Learning to Improve Cost and Duration Prediction Accuracy in Green Building Projects. *Journal of Construction Engineering and Management*,(2023) Vol.149(8).