International Journal for Housing Science and Its Applications
Publish August 3, 2025. Volume 46, Issue 3 Pages 258-270

https://doi.org/10.70517/ijhsa46318

Improvement of the effect of BIM education system innovation based on support vector machine on the cultivation of intelligent construction professional talents

Qizhi Zou[1], Qian Wu[2], Yuling Cui[1] and Xili Xie[1,*]
[1] School of Architectural Engineering, Qingdao Hengxing University of Science and Technology, Qingdao, Shandong, 266100, China
[2] Sichuan Tuolizhicheng Education Science and Technology Research Institute, Chengdu, Sichuan, 610041, China
Corresponding authors: (e-mail: xiexl@qdhxu.edu.cn).

Abstract With the rapid development of the intelligent construction industry, the demand of enterprises for high-quality talents specialized in intelligent construction is becoming more and more urgent. This paper discusses the evaluation model of talent cultivation effect of intelligent construction professionals under the innovation of BIM education system based on support vector machine (SVM) and principal component analysis (PCA). Firstly, the evaluation index system of intelligent construction professional talent training effect is constructed, due to the excessive evaluation indexes of intelligent construction professional talent training effect, the multi-dimensional data of talent training effect is downgraded by principal component analysis to avoid the estimation distortion of the model. Then the principal components with a determined number are transformed into the principal component matrix and input into the SVM, and the parameters of the SVM are selected based on the grid search algorithm, and the SVM is trained. The study shows that the assessment model of this paper can effectively evaluate the training effect of intelligent construction professionals in different institutions, and obtains the average absolute error E_{MA} and the mean square error E_{MS} with the optimal performances, and the values of the two indexes are 0.03 and 0.005, respectively. The innovative education effect assessment method of this paper provides a scientific basis for the optimization of the teaching content and the improvement of teaching methods. It provides a scientific basis and lays a foundation for improving the quality and knowledge level of intelligent construction professionals.

Index Terms support vector machine, grid search algorithm, principal component analysis, talent training effect

I. Introduction

In recent years, the construction industry has been facing the real pressure of high enterprise construction costs, low construction efficiency, low industrialization, high resource consumption, and shortage of high-quality labor [1]-[3]. At present, in the environment of artificial intelligence and other advanced technologies to promote the continuous progress of productivity in various industries, the construction industry is urgently required to take new technologies such as intelligent construction as an important kinetic energy to drive the development of the industry, centering on the overall goal of the high-quality development of the construction industry, taking intelligent construction as a hand to further strengthen the application of new technologies in the whole life cycle of construction projects, promoting the transformation of the construction industry to digitalization, and forming a system that covers scientific research, design, production, construction, operation and other whole industry chain integration of one industry system, highlighting the characteristics of the new quality productivity of the construction industry [4]-[8]. The requirements of the construction industry for the position is to master a construction skills based on the ability to understand the assembly industrialization process, building intelligence technology, building information modeling (BIM) technology applications and other emerging areas of technology, to efficiently solve the design, testing, construction, acceptance and delivery of the actual business problems faced by the enterprise [9]-[11]. In line with the development trend of industrialization, intelligence, and digitization in the construction industry, the intelligent construction profession needs to upgrade the talent training model currently being implemented, incorporate new technologies such as intelligent construction, green low-carbon, and digital integration, and provide high-quality technical and skilled personnel to support the development of the entire construction industry [12]-[15]. However, the resistance is heavy and the progress is slow, which is mainly manifested in the fact that the integration of digital technology, artificial intelligence and other modern information technology is not deep enough, especially the shortage of high-quality composite talents. In recent years, many colleges and universities have added new intelligent construction specialties, as a new specialty, how to train qualified personnel has become a difficult problem in the education sector and must be resolved.

BIM technology is another key technology in the informatization process of the civil construction industry following the traditional computer-aided design (CAD) software, which brings an innovative model to the engineering and construction field. BIM technology is different from traditional CAD technology, which has the characteristics of visualization, optimization, comprehensive modeling simulation scene, providing students with a closer experience of actual construction projects, and greatly stimulate the practical ability of students as well as innovation and entrepreneurship, and gradually realize the integration of industry-university-research education [16]-[19]. BIM, as the foundation of intelligent construction, is a form of education for students to practice technology-data-problems about BIM, and its teaching effect directly affects the quality of intelligent construction personnel training. However, the current teaching methods and effect evaluation system of BIM education present homogenization, homogenization, fragmentation, and lagging, resulting in insufficient integration of industry and education and poor practical skills of students [20]-[23]. The BIM education system needs to be innovated in order to better promote the cultivation of intelligent construction professionals, which has an important role in promoting the quality of education, which is not only the enhancement of education itself, but also a positive response to the diversified employment and entrepreneurial needs of the new era of economic and social development.

The article takes intelligent construction professionals from different institutions as the research object, and in response to the problem of too many evaluation indexes of talent cultivation effect and small samples, the principal component analysis (PCA) algorithm is used to downsize a large amount of evaluation index data, which not only reduces the dimensionality of the input features, but also eliminates the problems caused by the correlation between the data. The principal components are determined by calculating the principal components and the cumulative variance contribution rate, and the principal components are input into the support vector machine to complete the construction of the PCA-SVM talent cultivation effect evaluation model, and the first-level indicators and nine second-level indicators in the three aspects of safeguard support, cultivation process, and cultivation effect are determined on the basis of the collation and statistics of high-frequency indicators. Subsequently, the rationality of this paper's assessment model to innovate the BIM education system is verified through comparative experiments and example analysis.

II. Innovation of BIM education system based on support vector machine

II. A.SVM algorithm

Support Vector Machine (SVM) [24] is a machine learning method based on statistical learning theory.The goal of SVM is to train on given positive and negative samples to get an optimal classifier that is able to classify new samples.

Given the training samples $x_1, x_2, ..., x_N \in R^n$, x_1 to x_N denote their feature vectors, where N denotes the number of samples, and n denotes the feature dimensions of the samples. For N samples, each sample x_i corresponds to a category y_i, then there are $y_1, y_2, ...y_N \in \{-1, +1\}$, where +1 denotes the category of positive samples and -1 denotes the category of negative samples. The decision hyperplane is defined as shown below.

$$w \cdot x + b = 0 \tag{1}$$

To find the optimal dividing surface, it is sufficient to determine the weight vector w and the constant b. The marginal hyperplanes on either side of the dividing hyperplane are called support surfaces, and the points on the support surfaces are called support points (support vectors). Therefore, the weight vector w and the constant b can be determined from these support points and the expression for the decision hyperplane is shown in equation (2).

$$g(x) = (w, x) + b \tag{2}$$

Assuming the sample point $P(x_i, y_i)$, then according to the formula for the distance between a point and a plane, the distance from the sample point to the plane can be calculated as shown in equation (3).

$$d_i = \frac{y_i(w \cdot x_i + b)}{w} \tag{3}$$

Then, the minimum interval $d = \min(d_i)$ based on this hyperplane is found as the distance from the support vector to the hyperplane. The minimum interval is shown in equation (4).

$$d = \frac{1}{2}\left(\frac{2}{\| w \|}\right) \tag{4}$$

The SVM search for the maximum margin can then be expressed as an optimization problem with constraints, see Equation (5).

$$\begin{cases} \max_{w,b}(d) \\ subject\ to\ \dfrac{y_i(w\cdot x_i+b)}{\|w\|}\geq d, i=1,2,3,\ldots N \end{cases} \tag{5}$$

Since d and $\|w\|$ are scalars and maximizing $1/\|w\|$ is equivalent to minimizing $\|w\|\|^2/2$, the simplification leads to equation (6).

$$\begin{cases} \min\limits_{w,b}(\frac{1}{2}\|w\|^2) \\ subject\ to\ y_i(w\cdot x_i+b)\geq 1, i=1,2,3,\ldots N \end{cases} \tag{6}$$

At this point, the classification problem is also transformed into a constrained minimization problem which is constrained to satisfy $g(x)\leq -1$ for all negative samples and $g(x)\geq 1$ for all positive samples, which can be expressed as the optimization problem:

$$\begin{cases} \min\limits_{w,b}(\frac{1}{2}\|w\|^2) \\ y_i(w\cdot x_i+b)\geq 1, i=1,2,3,\ldots N \end{cases} \tag{7}$$

Replacing constraints with penalty terms converts a constrained optimization problem into an unconstrained optimization problem.

$$\min_{w,b}(\frac{1}{2}\|w\|^2)+\text{Penalty term} \tag{8}$$

For each data sample in the training data, define the penalty term as follows:

$$\left.\begin{cases} 0 & y_i(w_{x_i}^T+b)\geq 1 \\ \infty & \text{other} \end{cases}\right\} = \max_{\alpha_i\geq 0}\alpha_i(1-y_i(w^Tx_i+b)) \tag{9}$$

The optimization problem for rewriting the SVM is equation (10).

$$\begin{aligned} &\min\left(\frac{1}{2}\|w\|^2+\sum_{i=1}^{n}\max_{\alpha_i\geq 0}\alpha_i(1-y_i(w^Tx_i+b))\right) \\ &=\min_{w,b}\max_{\{\alpha_i\geq 0\}}\left\{\frac{1}{2}\|w\|^2+\sum_{i=1}^{n}\alpha_i(1-y_i(w^Tx_i+b))\right\} \end{aligned} \tag{10}$$

To make the above problem transform into a dyadic problem for an optimal solution, the KKT condition also needs to be satisfied.

$$\begin{cases} \alpha_i\geq 0 \\ y_i(w\cdot x_i+b)\geq 1 \\ \alpha_i(y_i(w\cdot x_i+b)-1)=0 \end{cases} \tag{11}$$

After the training is completed, most of the training samples do not need to be retained, and the final model is only related to the support vectors, which is called the sparsity of the solution of the support vector machine.

Suppose we need to classify and predict the intelligent construction professional talent training effect data x, the first step is to substitute the predicted sample x into $f(x)=w^Tx_i+b$, calculate the result of $f(x)$, and judge its category according to the plus or minus sign of its result. The final classification function is obtained as follows.

$$f(x)=\left(\sum_{i=1}^{n}\alpha_iy_ix_i\right)^T x+b=\sum_{i=1}^{n}\alpha_iy_i<x_ix>+b \tag{12}$$

Since in the pairwise problem of linear support vector machine learning, both the objective function and the classification decision function [25] involve only instances and inner products between instances, instead of displaying the specified nonlinear transformations, the kernel function is used to replace the inner products therein. The kernel function $K(x_i, x_j)$ is a function (or positive definite kernel) that implies the existence of a mapping $\phi(x)$ from the input space to the feature space, with x_i and x_j for any input space:

$$K(x_i, x_j) = \phi(x_i)\phi(x_j) \tag{13}$$

Considering general nonlinear classification problems, SVMs can be converted to optimization problems by soft-interval optimization and inner product kernel tricks.

$$\begin{cases} \min \dfrac{1}{2} \| w \|^2 + C\sum_{i=1}^{m} \varepsilon_i \\ y_i(w^T \cdot \phi(x_i) + b) \geq 1 - \varepsilon_i \\ \varepsilon_i \geq 0 \end{cases} \tag{14}$$

II. B.Principles of PCA data dimensionality reduction

Principal Component Analysis (PCA) [26] is a commonly used data dimensionality reduction method, which transforms high-dimensional data to a lower space by a linear method, extracts the main feature components of the data, and retains the key information in the original data. These key information are called principal components and can represent the vast majority of the information of the original variables. Since principal components are linear combinations of the original information, each principal component is linearly independent of each other, in other words, the use of PCA in data analysis avoids model estimation distortions or difficulties in estimating accurately due to the existence of precise or highly correlated relationships between variables.

Since this method is widely used in performing image recognition as well as high-dimensional data dimensionality reduction processing, it can not only remove the useless noise, but also reduce the amount of computation. Therefore, this paper considers the combination of PCA and SVM techniques.

According to the knowledge of matrix theory, a set of vectors can be accurately described by determining a set of bases and the projection of the vectors on the line where the bases are located. The same set of vectors can be represented differently by choosing different bases. When the number of bases used is smaller than the dimension of the vector itself, the vector can be downscaled.

Suppose there exists a vector set of dimension N, if its dimension is to be reduced from n dimensions to k dimensions $(k < n)$ and the original information of the vector set is to be retained to the maximum extent possible, it is necessary to choose k bases such that the projection of the vector set in the direction of this set of bases is as dispersed as possible.

Mathematically, it is common to use variance to describe the degree of dispersion of a set of data. A larger variance indicates that the elements in the set of data are more dispersed. For a vector group A consisting of m n-dimensional vectors, the variance is the mean of the sum of the squares of the differences between each vector a_i in the vector group and the mean value μ of all the vectors in the vector group, as follows:

$$Var(A) = \frac{1}{m}\sum_{i=1}^{m}(a_i - \mu)^2 \tag{15}$$

By zero-averaging each vector in the vector set, the variance of the vector set can be represented by the following equation:

$$Var(A) = \frac{1}{m}\sum_{i=1}^{m}a_i^2 \tag{16}$$

The above problem can be transformed into finding a set of bases that can maximize the value of the variance of all vectors when represented by them. In addition, for vector groups of high dimensionality, the correlation between individual vectors in the vector group needs to be measured by computing the covariance of the vector group. For a vector group A, its covariance is the mean of the sum of the differences of the two vectors a and b in the vector group multiplied by their respective means μ_a and μ_b, as shown in the following equation:

$$Cov(a,b) = \frac{1}{m-1}\sum_{i=1}^{m}(a_i - \mu_a)(b_i - \mu_b) \tag{17}$$

If all vectors in a vector group are zero-averaged, the covariance of that vector group can be represented by the following equation:

$$Cov(a,b)=\frac{1}{m}\sum_{i=1}^{m}a_i b_i \tag{18}$$

Two vectors are linearly independent when their covariance is zero. In order for the chosen basis to be able to represent more original information without repetition, it is desired that the bases be linearly independent of each other. The above problem can then be transformed into finding k unit orthogonal bases such that the covariance of a vector set transformation when represented by it is 0 and the variance of the variables is maximized.

Suppose there are two vectors $\vec{a}=(a_1,a_2,\cdots,a_m)$ and $\vec{b}=(b_1,b_2,\cdots,b_m)$ and they are formed into a matrix X by rows as follows:

$$X=\begin{pmatrix} a_1 & a_2 & \cdots & a_m \\ b_1 & b_2 & \cdots & b_m \end{pmatrix} \tag{19}$$

Compute its covariance matrix C for matrix X as follows, the diagonal elements of matrix C are the variances of vector $\vec{a}$ and vector $\vec{b}$, while the other elements are the covariances of vector $\vec{a}$ and vector $\vec{b}$.

$$C=\frac{1}{m}XX^T=\begin{pmatrix} \frac{1}{m}\sum_{i=1}^{m}a_i^2 & \frac{1}{m}\sum_{i=1}^{m}a_i b_i \\ \frac{1}{m}\sum_{i=1}^{m}a_i b_i & \frac{1}{m}\sum_{i=1}^{m}b_i^2 \end{pmatrix}=\begin{pmatrix} Cov(a,a) & Cov(a,b) \\ Cov(b,a) & Cov(b,b) \end{pmatrix} \tag{20}$$

Extending this to the general case, suppose that there are m vectors of dimension n, which are arranged row-wise into a matrix X, and let $C=\frac{1}{m}XX^T$, then the matrix C is a symmetric matrix whose diagonal corresponds to the variances of the individual variables, and the i th row, j th column, and the j th row, i th column, have the same elements, denoting the two covariance of the two variables.

In order to satisfy the dimensionality reduction condition that the data are as dispersed and uncorrelated as possible, it is necessary to find a set of bases such that the covariance matrix after transforming the matrix X to this set of bases has zero elements except for the diagonal, and the elements on the diagonal are arranged in order of magnitude.

Assuming that matrix P is a matrix consisting of a set of bases by rows, and matrix Y is the matrix after transforming matrix X to the set of bases, as shown in the following equation:

$$Y=PX \tag{21}$$

Then the covariance matrix D of matrix Y, as shown in the following equation:

$$\begin{aligned} D&=\frac{1}{m}YY^T \\ &=\frac{1}{m}(PX)(PX)^T \\ &=\frac{1}{m}PXX^TP^T \\ &=P(\frac{1}{m}XX^T)P^T \end{aligned} \tag{22}$$

If the matrix D satisfies that all elements except the diagonal are 0 and the elements on the diagonal are to be arranged in order of size, then the matrix PCP^T is a diagonal matrix and the elements on the diagonal are to be arranged in descending order. Since matrix C is a real symmetric matrix, one can find n unit-orthogonal eigenvectors $e_1,e_2,\ldots,e_n$, and form the matrix $E=(e_1,e_2,\cdots,e_n)$ by columns, and for matrix C can be obtained:

$$E^T CE = \Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} \tag{23}$$

where the matrix Λ is a diagonal matrix and $\lambda_1, \lambda_2, L, \lambda_n$ are the eigenvalues of the matrix C.

Therefore, let matrix $P = E^T$, i.e., the eigenvectors of covariance matrix C are obtained and unitized, and then the eigenvectors are sorted in order of eigenvalue magnitude in the diagonal matrix Λ to form matrix P, and then the matrix formed by selecting the first k eigenvectors of matrix P and multiplying them with the original matrix can be obtained to obtain the reduced dimensionality matrix to meet the requirements. The matrix can be obtained by multiplying the first k eigenvectors of matrix P with the original matrix.

II. C.PCA-SVM Talent Cultivation Effectiveness Evaluation Model

According to the actual situation, after processing and analyzing the intelligent construction professional talent training effect evaluation data, this paper proposes a PCA-SVM-based intelligent construction professional talent training effect evaluation model. First, the input features are downscaled using the PCA algorithm to eliminate the correlation between the factors in the original data, and then the determined principal component matrix data are used as the input of the SVM model, so as to optimize the training effect of the SVM model. Figure 1 shows the process of PCA-SVM model building.

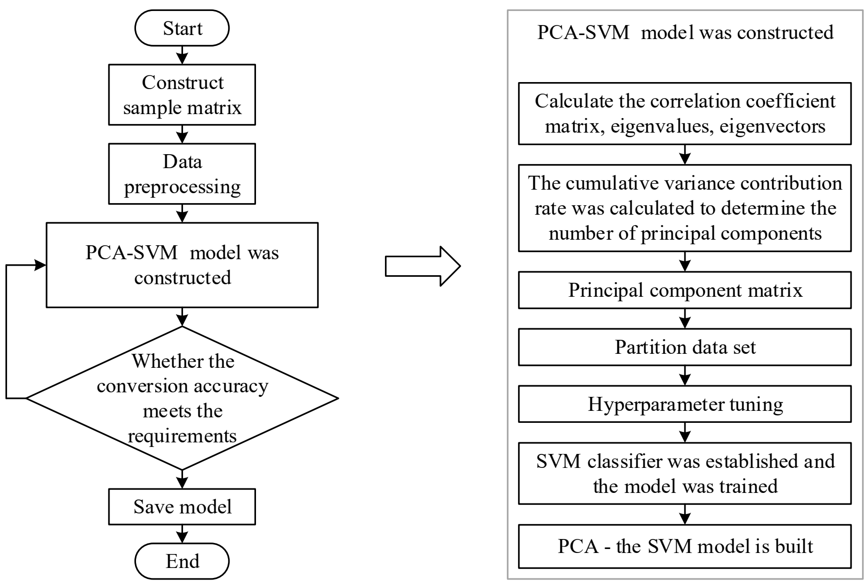

Figure 1: The process of the PCA-SVM model

The specific establishment steps of the intelligent construction professional talent training effect evaluation model based on PCA-SVM are as follows:

Step 1: Construct the evaluation dataset of intelligent construction professional talent cultivation effect. Collect intelligent construction professional talent cultivation data and select m important indicators affecting talent cultivation effect according to the principle of evaluating indicator screening, and take the evaluation results of talent cultivation effect as labels, so as to construct the input feature X and output label matrix Y.

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \tag{24}$$

Step 2: Data preprocessing. Data cleaning work is carried out on the intelligent construction professional talent training effect evaluation dataset. Firstly, the integrity and validity of the data are ensured by means of smoothing the noisy data, filling in the missing values and deleting the outliers. Then the data label Y is encoded using Label

Encoder to map the discrete data format of the original category label Y into a numerical coded representation that can be used for model training, and finally the data is standardized using Z-Score normalization operation to eliminate the differences in the magnitude and range of values.

Step 3: Calculate the covariance matrix X_{nom} for the normalized matrix XX^T and perform eigenvalue decomposition of the matrix XX^T to calculate the eigenvectors corresponding to the eigenvalues, and standardize all the eigenvectors to form the eigenvector matrix W.

Step 4: Calculate the interpretable variance contribution rate and cumulative variance contribution rate.

$$t_i = \frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i} \tag{25}$$

where t_i is explainable variance contribution and λ_i is size of interpretable variance.

$$E = \sum_{i=1}^{t} t_i \tag{26}$$

where E is cumulative variance contribution ratio.

Step 5: Determine the principal components. Determine the number of principal components by selecting the b principal components with a cumulative variance contribution rate of 95% or more, so as to reduce the dimensionality of the original data from the m dimension to the b dimension, and then construct the reduced dimension sample matrix D.

Step 6: Divide the data set. Divide the intelligent construction professional talent training effect evaluation dataset into training set and test set.

Step 7: Model hyperparameter tuning. Determine the optimal parameter combination of SVM classifier using grid search algorithm.

Step 8: Training model. The optimal parameter settings determined in Step 7 are then used to train the SVM classifier using the training set data, and the model will find the optimal hyperplane to classify the samples of different categories according to the features and labels of the samples.

Step 9: Observe whether the model accuracy meets the requirements. If it meets, go to the next step, otherwise adjust the model parameters and retrain until the model accuracy no longer improves.

Step 10: Save the model weights with the best training effect, use the test set to evaluate the generalization ability of the model, and finally obtain the PCA-SVM-based intelligent construction professional talent training effect evaluation model.

II. D.Establishment of Talent Cultivation Effect Evaluation Indicator System

The process of constructing the evaluation index system of talent cultivation effect for intelligent construction majors in this study is mainly divided into three steps. The first step is to search for journals, master and doctoral theses, research reports, policies and regulations that are highly relevant to the evaluation of teaching quality and talent assessment and evaluation of intelligent construction majors, and to statistically identify the high-frequency indicators.

Table 1: The quality evaluation system of talent cultivation

Primary indicator	Secondary indicator
Support system(X)	Human support(X1)
	Material support(X2)
	Supervisory management(X3)
Culture process(Y)	Training system(Y1)
	Competition system(Y21)
	Chemical acquisition(Y3)
Culture effect(Z)	Competition results(Z1)
	Student rank(Z2)
	Student delivery(Z3)

In the second step, the obtained high-frequency indicators are organized and classified. Based on the division of intelligent construction professional talent cultivation effect, the indicators are subdivided according to their interrelationships. Subsequently, on the basis of consulting relevant experts' opinions, the draft evaluation index system is constructed with the principles of scientificity, operability and systematicity as guidelines.

In the third step, the Delphi method was used to consult the experts' opinions on the draft evaluation indicator system, and the indicator system was adjusted and modified according to the expert group's opinions in order to determine the final evaluation indicator system. According to the two rounds of expert questionnaires and the screening of the screening basis, the evaluation index system for the evaluation effect of intelligent construction professional talent training after the innovation of BIM education system with 3 first-level indexes and 9 second-level indexes is finally established, as shown in Table 1.

III. Assessment of talent cultivation effect based on PCA-SVM

III. A. Talent Development Effectiveness Assessment Data Set

The expert group conducted an on-site assessment of the talent cultivation effect of intelligent construction majors in 50 institutions, in which some of the assessment data are shown in Table 2. The expert group conducts on-site assessment, usually by first setting up a diversified expert group consisting of school administrators, teachers, employer representatives, graduates, etc., and adopts questionnaires, on-site view assessment, data analysis of the school's talent cultivation status, and employer surveys, etc. to conduct the assessment. Institutions to be assessed are evaluated and assigned points (out of 10) for each of the 9 key assessment indicators in the Talent Cultivation Effectiveness Assessment Indicator System, and then the total score is calculated according to the weights of each indicator item, and the comprehensive assessment data is clearly divided into the classification levels of talent cultivation effectiveness, such as “excellent, good, and general”. The PCA-SVM method of assessing the effectiveness of talent training does not require the use of the weights of the indicators, nor does it calculate the total score.

Table 2: Assessment data division score

School number	X1	X2	X3	Y1	Y2	Y3	Z1	Z2	Z3
1	8.2	8.7	8.3	8.9	7.2	7.1	8.8	8.4	8.1
2	7.1	6.5	8.1	6.0	5.6	5.6	7.7	7.6	8.0
3	9.0	9.5	9.6	9.1	8.9	8.3	9.8	9.5	8.7
4	7.9	8.1	8.1	9.0	6.9	7.0	8.8	8.4	7.7
5	7.1	6.6	9.4	6.1	6.3	6.0	7.9	7.5	8.9
6	7.0	6.6	8.2	6.0	6.1	6.0	7.8	7.3	7.9
7	7.9	8.3	8.3	8.9	9.2	9.1	8.6	8.1	7.9
8	8.0	8.8	8.2	9.1	9.2	9.1	8.9	9.6	7.7
9	8.3	7.0	7.8	8.0	8.2	7.1	7.0	7.7	7.9
10	6.6	7.1	7.7	6.8	8.4	3.9	6.4	6.7	5.6

III. B. Downgrading of Talent Training Effectiveness Assessment Data

Apply PCA analysis to realize data dimensionality reduction. A new practice logic library was created in the SAS statistical analysis system, which was imported into the school talent cultivation effect assessment dataset zypg.sas7bdat, and the princomp function was applied to carry out the principal component analysis.

Table 3: Eigenvalue and cumulative contribution

Serial number	Eigenvalue	Adjacent eigenvalue difference	Contribution rate	Cumulative contribution
1	16.1210	13.5221	0.7014	0.7014
2	2.5989	0.9490	0.1138	0.8152
3	1.6497	0.3872	0.0809	0.8961
4	1.2622	0.8069	0.0546	0.9507
5	0.4545	0.1941	0.0190	0.9697
6	0.2609	0.0510	0.0108	0.9805
7	0.2087	0.0507	0.0088	0.9893
8	0.1580	0.0503	0.0064	0.9957
9	0.1068	0.0473	0.0043	1.0000

The eigenvalues of the correlation coefficient matrix and its cumulative contribution are shown in Table 3. Where the larger the eigenvalue, the more information is contained in its corresponding principal component variable. The first four items of the cumulative contribution rate have reached 0.9507, so these four principal components (i.e., the first four items of the new dataset generated by the principal component analysis, pc1, pc2, pc3, and pc4) are selected to replace the nine indicators in the original talent cultivation effect assessment data to realize the dimensionality reduction of the dataset. The dimensionality reduction of the assessment data of talent training effectiveness of some institutions is shown in Table 4.

Table 4: Four main component data after the descending dimension

School number	pc1	pc2	pc3	pc4
1	0.4430	0.2547	0.0617	0.4432
2	0.3522	0.1090	0.2625	0.3521
3	1.0000	0.9357	0.1497	0.9999
4	0.4437	0.2549	0.0617	0.4442
5	0.5830	0.8237	0.0009	0.5829
6	0.3509	0.1147	0.2694	0.3497
7	0.2389	0.5383	0.4203	0.2384
8	0.0854	0.9941	0.1155	0.0846
9	0.2385	0.5390	0.4199	0.2390
10	0.0850	0.9947	0.1162	0.0849

III. C. Feature model training and testing for school grades

The professional rating feature model training and testing were carried out by applying Libsvm software, which is a SVM pattern recognition and regression software package characterized by relatively few parameter adjustments involved in SVM and more default parameters. In order to ensure the training effect of the model and avoid the situation that small features are covered by large features due to the large level difference of the scoring value of each index item, the professional assessment data are normalized in the interval of [0, 1] before starting the training of the feature model.

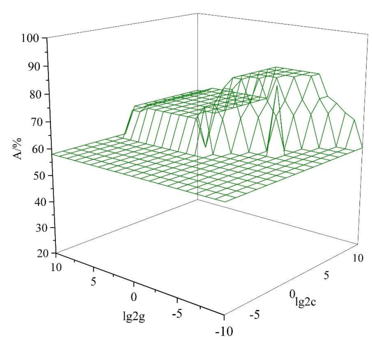

Figure 2: Downdimension preparameter optimization result

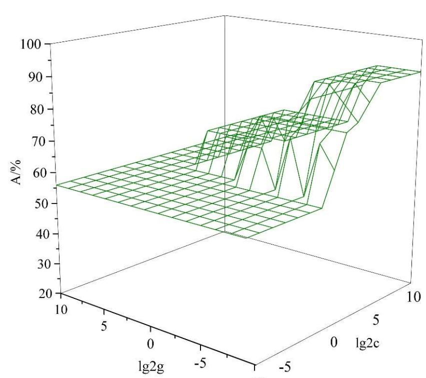

Figure 3: Downway parameter optimization result

Parameter selection in Libsvm software. Call the lattice parameter optimization function SVMcgForClass to achieve c, g parameter optimization, the results of the optimization in this case are shown in Figure 2 and Figure 3. Among them, the optimal parameters of the talent cultivation effect evaluation data set before dimensionality reduction are c=4, g=0.0153. The optimal parameters of the corresponding talent cultivation effect evaluation data set after dimensional reduction are c=4, g=1.

Next, model training and testing are performed, and the best parameters and classification accuracy of professional assessment before and after dimensionality reduction are shown in Table 5.

(1) Model training and testing on the data before dimensionality reduction.

The 25 assessment data from the above dataset are selected as the training sample set, and the remaining 25 professional assessment data are used as the testing sample set. Then, the Svmstrain model training function and the Svmpredict test function were invoked respectively, and the classification accuracy obtained was 91.1452%. The results show that the classification results obtained by applying Libsvm software coincide with the construction results derived from the expert group assessment at 91.1452%.

(2) Model training and testing of the dimensionality reduced data. The corresponding principal component data after dimensionality reduction in the evaluation data are selected, and then the Svmstrain model training function and the Svmpredict test function are called respectively, and the classification accuracy rate obtained is 95.6879%. This accuracy is 4.5427% higher than that before dimensionality reduction. Through data dimensionality reduction, the overlapping part of information in the samples is removed, and the classification accuracy is improved.The agreement rate between the PCA-SVM evaluation method and the on-site evaluation results of the expert group is as high as 95.6879%, which indicates the reliability of the PCA-SVM evaluation method and can replace the on-site evaluation work of the expert group.

Table 5: The accuracy of classification is accurate

Libsvm classification accuracy %		Optimum parameter	
Before the descending dimension	After the descending dimension	c	g
65.2145	95.6879	4	1
91.1452	52.1458	4	0.0153

III. D. PCA-SVM fitting effects

III. D. 1) Experimental design

Aiming at the shortcomings of using a single support vector machine to analyze the slope safety coefficient that there is correlation between the input variables and too much input data, so this paper combines the principal component analysis method with the support vector machine to construct a preferred model for a reasonable analysis. From the actual assessment results, 950 sets of data are randomly selected for model training, and 50 sets of data are selected for model validation, where the 10-fold crossover method is used for validation, and the mean absolute error E_{MA} and mean square error E_{MS} are used as evaluation indexes.

III. D. 2) Comparative Experimental Results

The comparison results of evaluation indexes and the data fitting effect are shown in Table 6 and Figure 4 respectively. Compared with the SVM model, the fitting effect of this paper's model is better, and the fitted curve in the evaluation of the training effect of intelligent construction professionals is closer to the real evaluation curve, and the two evaluation indexes of E_{MA} and E_{MS} are reduced by 85.71% and 94.12%, respectively, compared with the SVM model.

Table 6: Comparison of the evaluation indicators of PCA-SVM and SVM

	PCA-SVM	SVM
E_{MA}	0.04	0.28
E_{MS}	0.01	0.17

The above experiments have shown that the use of PCA is effective, in order to verify the performance of PCA-SVM's compared to other models, the model in this paper is then compared to the models of the popular Genetic Algorithm Optimized BP Neural Networks (GA-BP), Gradient Enhanced Regression (GBR), Particle Swarm Optimized BP Neural Networks (PSO-BP) and RBF Neural Networks (RBF) (all the models were subjected to principal component analysis) and the fitting results are shown in Figure 5. Similarly, the effects of the fitting results of the five models were compared using the mean absolute error E_{MA} and the mean square error E_{MS} as the evaluation indexes, and the results are shown in Table 7.

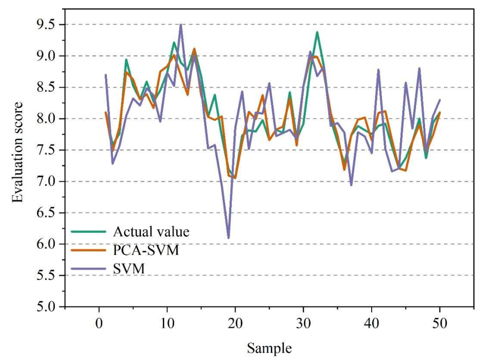

Figure 4: Data fitting rendering

The comparison graphs of the fitting curves of the five models show that the fitting effect of the model proposed in this paper is the closest to the real value, and it is the best fitting effect among all the models. The data in Table 7 also show that the error of the model proposed in this paper is the smallest, and the average absolute error E_{MA} and the mean square error E_{MS} are 0.03 and 0.005, respectively, which indicates that in the validation experiments, the PCA-SVM model has the highest stability, and it is suitable for evaluating the effect of talent training for intelligent construction professionals.

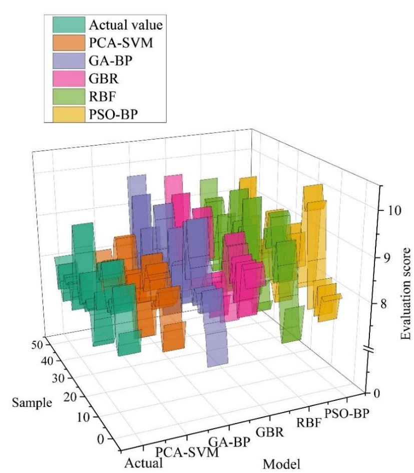

Figure 5: The comparison of the five model fitting results

Table 7: Comparison of evaluation indexes of 5 models

Model	E_{MS}	E_{MA}
GA-BP	0.075	0.23
PSO-BP	0.022	0.19
RBF	0.054	0.19
GBR	0.038	0.09
PCA-SVM	0.005	0.03

III. D. 3) Example analysis of experimental results

In this subsection, students of intelligent construction majors in four institutions numbered 20, 21, 22, and 23 located in the same urban area are selected as samples for talent training effect evaluation.The fitting effect of RBF and GBR is better, so GBR, RBF, and PCA-SVM are utilized to validate the above data, and the evaluation results are shown in Table 8.

The absolute errors of the PCA-SVM model for the prediction results of the talent cultivation effect of the four institutions are 0.004, 0.004, 0.012 and 0.004, respectively, which are the smallest compared with the other two models, and the predicted values are also closer to the actual values. Once again, it shows the effectiveness of the PCA-SVM model in the evaluation of the talent cultivation effect of intelligent construction majors.

Table 8: Comparison of evaluation results

Test school	20	21	22	23
RBF prediction	9.11	8.08	7.84	8.95
GBR prediction	9.24	8.04	7.89	8.86
PCA-SVM prediction	9.35	8.58	7.65	8.55
Actual value	9.39	8.55	7.56	8.45
RBF(MAE)	0.030	0.055	0.059	0.030
GBR(MAE)	0.016	0.060	0.049	0.016
PCA-SVM(MAE)	0.004	0.004	0.012	0.004

IV. Conclusion

In this paper, the combined method of principal component analysis and support vector machine technology was used to evaluate the cultivation effect of intelligent construction professional talents under the innovation of BIM education system. The dimensionality reduction processing of input data is realized by using PCA, which avoids the computational complication caused by the correlation between input variables, and the experiment shows that the classification accuracy obtained after dimensionality reduction is 95.6879%, which is 4.5427% higher than that before dimensionality reduction, so the application of Principal Component Analysis is effective for dimensionality reduction of complex data.

The PCA-SVM model proposed in this paper is compared with the commonly used genetic algorithm optimized BP neural network, gradient augmented regression, particle swarm optimized BP neural network and other models for evaluating the effect of intelligent construction professional talent training, and the fitting results of the PCA-SVM model are the closest to the evaluation scores of the actual talent training effect, and at the same time, the two evaluation scores of E_{MS} and E_{MA} two evaluation indexes are also the smallest, 0.005 and 0.03 respectively, which verifies the validity and feasibility of the assessment model in this paper, avoids a certain degree of assessment bias, and provides a new way of thinking for the research field of intelligent teaching effect assessment.

In addition to constructing a PCA-SVM-based evaluation model of talent cultivation effect, the innovation based on SVM and BIM education system can also be used to improve the effect of talent cultivation for intelligent construction majors through the establishment of a BIM practical teaching platform, the development of core items of practical training materials and other strategies. These strategies help students of intelligent construction majors to better master BIM technology in the learning stages of architectural design, construction management, etc., and provide assistance for students to improve their professionalism and knowledge ability level.

Funding

Shandong Province Senior Professors Association 2024 Annual Key Research Project on Aesthetic Education (Project Number: MYKTB0003).

References

[1] Shaikh, F. A., Odhano, N., & Kaliannan, S. (2020). Performance and management of cost in the construction industry. Civil Engineering Journal, 6(7), 1368-1374.

[2] Hu, X., & Liu, C. (2018). Measuring efficiency, effectiveness and overall performance in the Chinese construction industry. Engineering, Construction and Architectural Management, 25(6), 780-797.

[3] Yap, J. B. H., Chow, I. N., & Shavarebi, K. (2019). Criticality of construction industry problems in developing countries: Analyzing Malaysian projects. Journal of Management in Engineering, 35(5), 04019020.

[4] Blanco, J. L., Fuchs, S., Parsons, M., & Ribeirinho, M. J. (2018). Artificial intelligence: Construction technology's next frontier. Building Economist, The, (Sep 2018), 7-13.

[5] Guo, Z., & Li, L. (2022). A conceptual framework for collaborative development of intelligent construction and building industrialization. Frontiers in Environmental Science, 10, 904518.

[6] Naji, K. K., Gunduz, M., Alhenzab, F. H., Al-Hababi, H., & Al-Qahtani, A. H. (2024). A systematic review of the digital transformation of the building construction industry. IEEE Access, 12, 31461-31487.

[7] Lee, B., & Park, S. K. (2022). A Study on the Competitiveness for the Diffusion of Smart Technology of Construction Industry in the Era of 4th Industrial Revolution. Sustainability, 14(14), 8348.

[8] Wang, D., & Cheng, X. (2024). Study on the path of high-quality development of the construction industry and its applicability. Scientific Reports, 14(1), 14727.

[9] Abd Rahman Ahmad, K. S. Y., Isa, K., Soon, N. K., & Sapry, H. R. M. (2019). The job specification and job description for Indonesian workers in Malaysian construction industry. International Journal of Engineering and Advanced Technology (IJEAT), 8, 242-244.

[10] Li, J., Yang, M., Liu, C., Li, A., & Guo, B. (2023). Listen to the companies: exploring BIM job competency requirements by text mining of recruitment information in China. Journal of Construction Engineering and Management, 149(9), 04023076.

[11] Yu, J., Wang, J., Shi, Q., Xu, J., & Wang, J. (2024). Exploring job competency related to intelligent construction in China using a text mining method. Engineering, Construction and Architectural Management.

[12] Ying, D., & Li, N. (2023). Research on the Cultivation of Skilled Talent in Intelligent Construction to Facilitate a Model of Shared Prosperity. International Journal of Frontiers in Sociology, 5(16).

[13] Sh Said, S. A. A. (2023). Artificial intelligent (AI) in construction industry: the talent gap/Dr Sheikh Ali Azzran Sh Said. RISE: Catalysing Global Research Excellence, (3), 1-5.

[14] Wen, Y. (2021). Research on the intelligent construction of prefabricated building and personnel training based on BIM5D. Journal of Intelligent & Fuzzy Systems, 40(4), 8033-8041.

[15] Zhang, P., Ma, S. G., Zhao, Y. N., & Cao, X. Y. (2023). Analyzing core competencies and correlation paths of emerging engineering talent in the construction industry—an integrated ISM–MICMAC approach. Sustainability, 15(22), 16011.

[16] Govender, R., Saba, G., Ham, N., Hou, L., Moon, S., & Kim, J. J. (2022). Appraisal of building information modeling (BIM) curriculum for early-career construction-industry professionals: case study at C educational institute in Korea. International Journal of Construction Management, 22(5), 891-899.

[17] Liu, Z. (2024). Comparison and analysis of advantages and disadvantages between BIM and CAD in civil drafting software. Applied and Computational Engineering, 62, 192-197.

[18] Jolanta, Š., & Pupeikis, D. (2018). Review of BIM implementation in Higher Education. Journal of sustainable architecture and civil engineering, 22(1), 99-109.

[19] Sanchez-Lite, A., Gonzalez, C., Zulueta, P., & Sampaio, Z. (2020). A comparative study of the use of building information modeling in teaching engineering projects. IEEE Access, 8, 220046-220057.

[20] Shelbourn, M., Macdonald, J., McCuen, T., & Lee, S. (2017). Students' perceptions of BIM education in the higher education sector: A UK and US perspective. Industry and Higher Education, 31(5), 293-304.

[21] Huang, Y. (2018). A review of approaches and challenges of BIM education in construction management. Journal of Civil Engineering and Architecture, 12(6), 401-7.

[22] Chen, K., Lu, W., & Wang, J. (2020). University–industry collaboration for BIM education: Lessons learned from a case study. Industry and Higher Education, 34(6), 401-409.

[23] Li, J., Afsari, K., Li, N., Peng, J., Wu, Z., & Cui, H. (2020). A review for presenting building information modeling education and research in China. Journal of Cleaner Production, 259, 120885.

[24] Moy'awiah A Al Shannaq,Areen Nasrawi,Abed Al Raouf K Bsoul & Ahmad A Saifan. (2025). Abnormal heart sound recognition using SVM and LSTM models in real-time mode.. Scientific reports,15(1),9129.

[25] Wenjuan Huyan & Jing Li. (2021). Research on rural tourism service intellectualization based on neural network algorithm and optimal classification decision function. Journal of Ambient Intelligence and Humanized Computing,13(Suppl 1),1-21.

[26] Mohammed Tahar Habib Kaib,Abdelmalek Kouadri,Mohamed Faouzi Harkat,Abderazak Bensmail & Majdi Mansouri. (2025). Data size reduction approach for nonlinear process monitoring refinement using Kernel PCA technique. Expert Systems With Applications,274,126975-126975.