

Improved Kernel Limit Learning Machine Based on Vector Weighted Average Algorithm for Concrete Compressive Strength Prediction

Haoran Sun

*Chang' an Dublin International College of Transportation, Chang' an University, Xian, China
1552708626@qq.com*

Abstract: In this paper, a kernel-limit learning machine method based on an improved vector-weighted average algorithm is proposed for enhancing the accuracy of concrete compressive strength prediction. Quantitative relationships between components and compressive strength are revealed by Pearson correlation analysis: cement content shows the strongest positive correlation ($r=0.50$), followed by superplasticizer ($r=0.37$) and age ($r=0.33$), while water content shows a significant negative correlation ($r=-0.29$). The prediction model based on this model showed good performance in both the training and test sets, with a coefficient of determination (R^2) of 0.913 and a root mean square error (RMSEC) of 4.875 for the training set, and R^2 of 0.867 and RMSEP of 6.171 for the test set, It is worth noting that the model's prediction accuracy in the test set decreased by only 0.046 units of R^2 compared with the training set, and the error increase remained reasonable. and the error increase remains within a reasonable range, indicating that the improved algorithm has excellent generalization ability and engineering applicability. By integrating feature correlation and intelligent algorithm optimization, this study not only provides a new method for concrete material strength prediction, but also constructs a quantitative model that can provide an important theoretical support for the optimal design of concrete proportion and the assessment of structural durability of engineering structures, which is of practical application value for improving the efficiency of concrete material research and development and the level of engineering quality control.

Keywords: Vector weighted average algorithm, Kernel limit learning machine, Concrete compressive strength.

1. Introduction

The prediction of concrete compressive strength is one of the core topics in the field of construction engineering and materials science. As the most widely used building material in the world, the mechanical properties of concrete directly determine the safety, durability and economy of buildings. Traditionally, the determination of compressive strength mainly relies on destructive testing of standard laboratory curing specimens, which is a time-consuming and costly method that takes up to 28 days and is difficult to meet the demand for real-time quality control and construction efficiency in modern engineering [1]. In addition, the strength of concrete is subject to the nonlinear coupling of water-cement ratio, aggregate properties, additive types and curing conditions, etc., and its inner

mechanism is complex, so it is difficult for traditional empirical formulas or statistical models to accurately describe the complex relationship between the variables. With the advancement of green building and intelligent construction concepts, how to quickly and non-destructively predict the strength of concrete during the construction phase and reduce the waste of resources through ratio optimization has become a key issue to be solved by the industry. In this context, intelligent prediction methods integrating material science, data analysis and computer technology have gradually become a research hotspot [2].

Machine learning algorithms provide breakthrough solutions for concrete strength prediction. By mining the hidden laws in the historical experimental data, machine learning is able to establish a nonlinear mapping model between the input variables (e.g., cement content, fly ash admixture, age, etc.) and the compressive strength, overcoming the theoretical limitations of the traditional methods [3]. Algorithms such as Random Forest [4] and Support Vector Machine [5] can identify the key influencing factors through feature importance analysis; neural networks, especially deep learning models, exhibit higher prediction accuracy under complex proportioning conditions by virtue of the ability of multilayer nonlinear transformation [6]. In this paper, the kernel limit learning machine is improved based on the vector weighted average algorithm for prediction of compressive strength of concrete.

2. Data set sources and data analysis

We use the open source dataset for experiments, the source of the dataset is kaggle, the dataset contains the content and proportion of the basic components of concrete such as cement, blast furnace slag, fly ash, water, superplasticizers, coarse aggregates, fine aggregates, and age, and the target variable is the compressive strength of concrete, a total of 1,030 pieces of data, this paper selects some of the data to show, as shown in Table 1.

Table 1: Modelling assessment.

Cement	Blast Furnace Slag	Fly Ash	Water	Superplasticizer	Coarse Aggregate	Fine Aggregate	Age	Concrete compressive strength
540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.99
540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.89
332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.27
332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.05
198.6	132.4	0.0	192.0	0.0	978.4	825.5	360	44.30
266.0	114.0	0.0	228.0	0.0	932.0	670.0	90	47.03
380.0	95.0	0.0	228.0	0.0	932.0	594.0	365	43.70
380.0	95.0	0.0	228.0	0.0	932.0	594.0	28	36.45
266.0	114.0	0.0	228.0	0.0	932.0	670.0	28	45.85
475.0	0.0	0.0	228.0	0.0	932.0	594.0	28	39.29
198.6	132.4	0.0	192.0	0.0	978.4	825.5	90	38.07
198.6	132.4	0.0	192.0	0.0	978.4	825.5	28	28.02

It is very important to analyze the relationship between each basic component and the compressive strength of concrete, we use the method of Pearson correlation analysis to calculate the correlation between each cement component and the compressive strength of concrete, calculate the correlation coefficients between them and draw the correlation heat map, as shown in Figure 1, and sorted according to the degree of correlation with concrete, as shown in Figure 2.

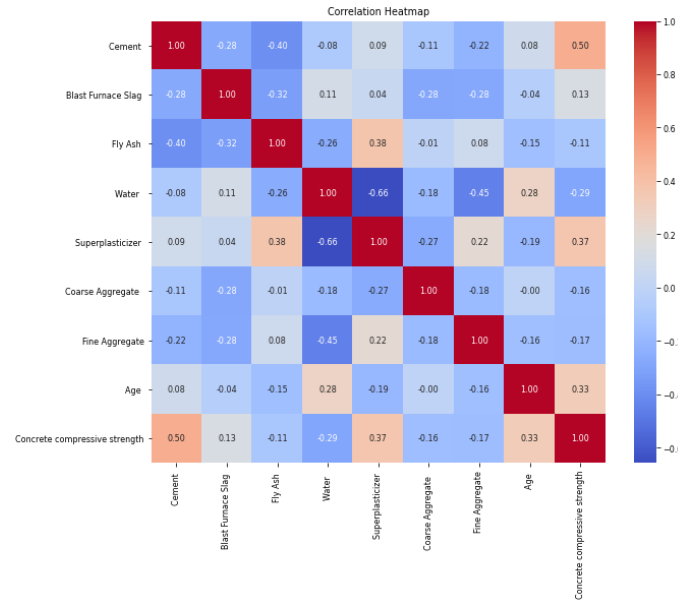


Figure 1: The correlation heat map.

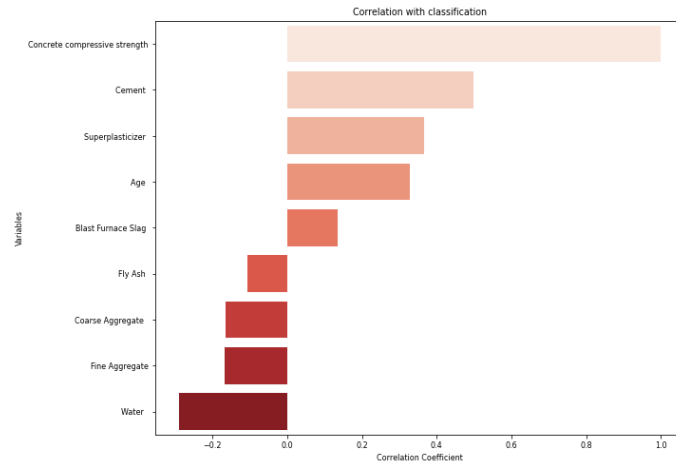


Figure 2: The degree of correlation with concrete.

From the correlation matrix, it can be seen that the most positively correlated variable with the compressive strength of concrete is cement, with a positive correlation coefficient of 0.5; followed by superplasticizer, with a correlation coefficient of 0.37, and age is also an important factor affecting the compressive strength of concrete, with a correlation coefficient of 0.33. It is worth noting that water has a negative correlation with the compressive strength of concrete, with a correlation coefficient of -0.29.

3. Method

3.1. Vector Weighted Average Algorithm

Vector weighted average algorithm is a data fusion method that integrates information from multiple vectors by assigning different weights, and its core idea is to assign specific weight coefficients to each vector according to the difference in importance or reliability of each vector, and then generate the integrated result vector by weighted summation [7]. In the calculation process, each input vector

will be multiplied with its corresponding weight, the larger the weight value represents the higher contribution of the vector to the final result, and after all the weighted vectors are superimposed, the sum of the weights is usually normalized to ensure the stability of the result. The key to this algorithm lies in the reasonable setting of the weights, which can either be specified manually based on a priori knowledge or automatically optimized through statistical analysis of data or machine learning models [8]. The pseudo-code of the vector weighted average algorithm is shown in Figure 3.

```

1)   Input the max. no. of generation ( $Maxg$ ), and population size ( $Np$ )
2)   Produce an initial population of  $Np$  solution vectors at random ( $P^0=\{x_1^0, x_2^0, \dots, x_{Np}^0\}$ )
3)   Evaluate each solution vector by the fitness function,  $f(x_i^0), i = 1, 2, \dots, Np$ 
4)   Obtain the best solution vector ( $x_{bs}$ )
5)   while  $g < Maxg$  do
6)     for  $i = 1$  to  $Np$  do
7)       Choose randomly  $a \neq b \neq c \neq i$  from the interval  $\{1, Np\}$ 
8)       Compute the solution vectors  $z1_i^g$  and  $z2_i^g$  by Eq. (21)      % Updating process
9)       Compute the solution vector  $u_i^g$  by Eq. (25)              % vector combination process
10)      Compute the local search phase by Eq. (26)              % Local search process
11)      Compute the OF value ( $f(u_i^g)$ )
12)      If  $f(u_i^g) < f(x_i^g)$  then  $x_i^{g+1} = u_i^g$ 
13)      Otherwise  $x_i^{g+1} = x_i^g$ 
14)    end for
15)    Update the best vector ( $x_{bs}$ )
16)     $g = g + 1$ 
17)  end while
18)  Return  $x_{bs}$  as the final solution

```

Figure 3: The pseudo-code of the vector weighted average algorithm.

3.2. Kernel limit learning machine

Kernel Extreme Learning Machine (KELM) is an efficient machine learning model that combines the Extreme Learning Machine (ELM) framework with the kernel method, aiming to improve the generalization ability and stability of traditional ELM through nonlinear mapping, while retaining its fast training characteristics. Traditional ELM avoids the time-consuming gradient descent process in traditional neural networks by randomly initializing the weights of a single hidden-layer neural network and solving the output weights directly analytically, but the randomness of its hidden-layer parameters may lead to fluctuations in the model's performance, and its generalization ability is limited especially under complex data distributions [9]. The core improvement of the KELM lies in the introduction of the kernel trick, which implicitly maps the original data to a high-dimensional feature space. thereby bypassing explicit random weight generation and constructing nonlinear decision boundaries directly by measuring the similarity between samples through the kernel function [10]. The schematic diagram of Kernel Extreme Learning Machine (KELM) is shown in Fig. 4.

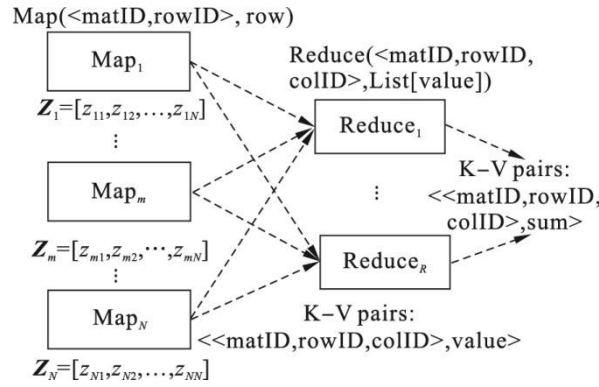


Figure 4: The pseudo-code of the vector weighted average algorithm.

3.3. Kernel limit learning machine optimized for vector weighted average algorithm

The performance of KELM is highly dependent on the selection of the kernel function and its parameters, for example, the bandwidth parameter of the Gaussian kernel determines the degree of local sensitivity of the feature space, which needs to be adjusted according to the characteristics of the data in order to adapt to the nonlinear structure. Compared with the traditional ELM, KELM eliminates the instability caused by the randomness of the hidden layer through kernel mapping, which significantly improves the robustness and consistency of the model, especially in the small-sample or high-dimensional nonlinear tasks.

The algorithm uses a two-level weight update strategy:

1. Sample-level weights: adjusted in real time according to the prediction errors of the samples during training. The samples with wrong prediction will be weighted down to prevent the model from overfitting the local noise; the weights of correctly predicted samples will be gradually increased to strengthen the model's capture of stable patterns.

2. Kernel function level weights: Fusing the outputs of multiple kernel functions (e.g., Gaussian kernel, linear kernel), the weights of different kernels are dynamically adjusted to balance the model's local fitting ability and global generalizability. For example, for data regions with fuzzy boundaries, the algorithm may increase the weight of the Gaussian kernel to capture details; for linearly separable regions, it focuses on the simplicity of the linear kernel.

4. Result

In the experimental setup, the Kernel Extreme Learning Machine (KELM) adopts the radial basis (RBF) kernel function, the regularization parameter C is set to 100 to balance the risk of overfitting, and the kernel parameter γ is set to 0.1 to control the kernel function bandwidth; the vector weighted averaging algorithm assigns dynamic weights to the training set by calculating the information entropy of the feature vectors in the training set, and the entropy weighting threshold is set to 0.7 to screen the effective features. The dataset was divided into training set and test set according to 7:3, and the optimization parameters were optimized using 5-fold cross-validation; the regression performance was evaluated by mean square error (MSE), mean absolute error (MAE) and coefficient of determination (R^2); all experiments were repeated independently for 10 times to take the mean; the data were preprocessed using Z-score standardization to eliminate the effect of magnitude.

The model was introduced for training, and a line graph of the actual values of the training set versus the predicted values of the model was output, with the actual concrete strength in red and the predicted values of our model in blue. The line plot of actual values versus predicted values of the

model for the training set is shown in Figure 5, and the line plot of actual values versus predicted values of the model for the test set is shown in Figure 6.

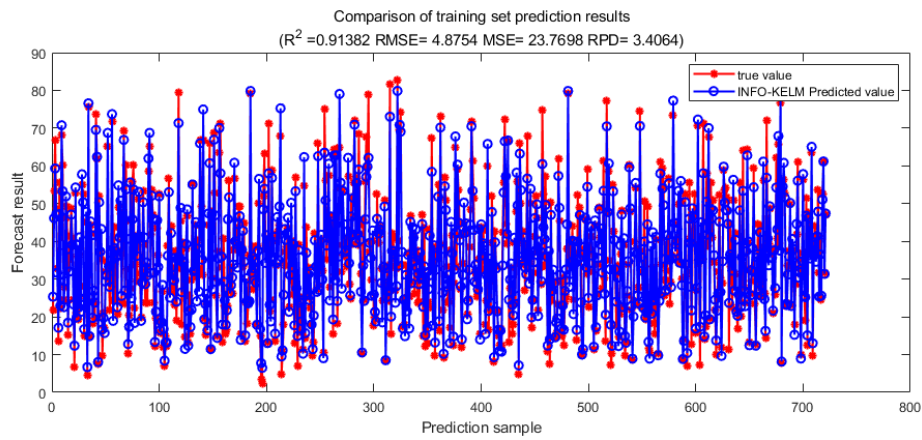


Figure 5: The line plot of actual values versus predicted values of the model for the training set.

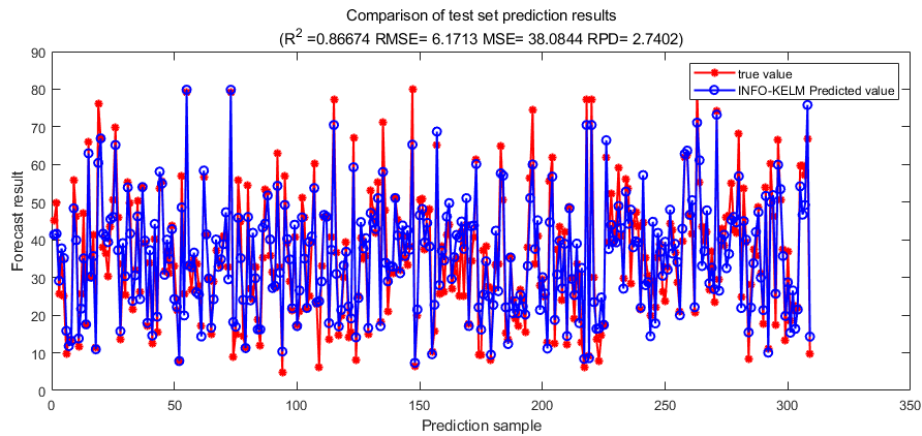


Figure 6: The line plot of actual values versus predicted values of the model for the test set.

From the prediction effect of the training set and test set, our model can predict the concrete compressive strength more accurately from the perspective of quantization.

Secondly, we use the quantitative method to observe more intuitively the prediction ability of our model for concrete strength on the training and test sets. The scatter plot of the actual values of the predicted values in the training set is shown in Fig. 7, and the scatter plot of the actual values of the predicted values in the test set is shown in Fig. 8.

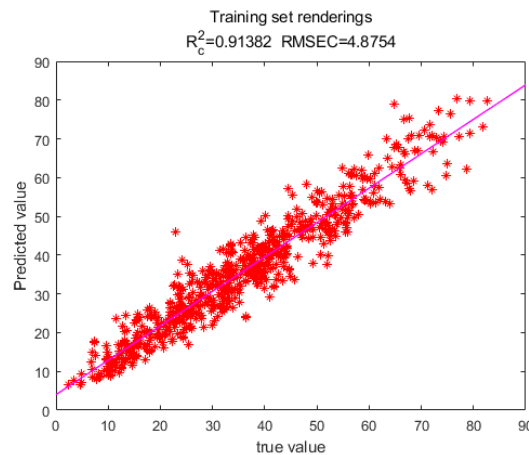


Figure 7: The line plot of actual values versus predicted values of the model for the test set.

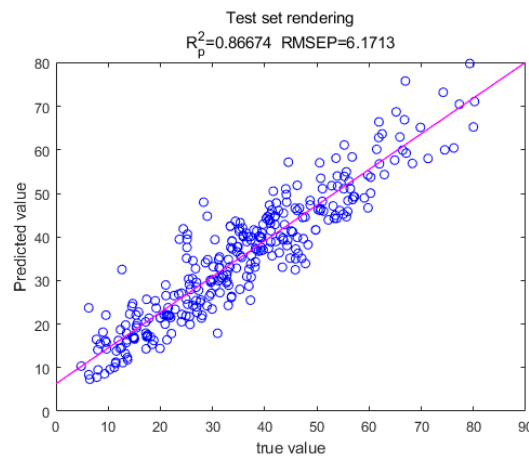


Figure 8: The line plot of actual values versus predicted values of the model for the test set.

From the scatter distributions of predicted-actual values in the training and test sets, it is clear that our model is able to predict the concrete compressive strength relatively accurately. From the quantitative point of view, the R^2 of the training set is 0.913 and the RMSEC is 4.875, and the R^2 of the test set is 0.867 and the RMSEP is 6.171. Among them, the test set is lower than the training set by 0.046 in R^2 , which indicates that the effect of the test set is reduced compared with the training set, but the value of the reduction is not large, which indicates that the model's ability to generalize is better.

5. Conclusion

This article uses the Newton Raphson algorithm to optimize the XGBoost model and applies it to advertising click prediction tasks. Through the analysis of experimental data, we found that there is a certain correlation between advertising click through rate and multiple variables. Among them, the variable most positively correlated with ad click through rate is ad position, with a correlation coefficient of 0.08; The variable most negatively correlated with ad click through rate is browsing history, with a correlation coefficient of -0.08. These correlation analyses provide important reference for the construction of the model. During the model training process, we observed that the fitness curve of the model showed a significant downward trend, gradually decreasing from the initial 0.184 to 0.148, indicating that the predictive performance of the model continued to improve with the

progress of training. This trend further validates the effectiveness of the Newton Raphson algorithm in optimizing the XGBoost model. In addition, this study also reveals the significant impact of advertising location and browsing history on ad click through rates. The position of advertisements is positively correlated with click through rates, indicating that the more prominent the advertisement is on the page, the greater the likelihood of users clicking.

References

- [1] El-Mir, Abdulkader, et al. "Machine learning prediction of concrete compressive strength using rebound hammer test." *Journal of Building Engineering* 64 (2023): 105538.
- [2] Zeng, Ziyue, et al. "Accurate prediction of concrete compressive strength based on explainable features using deep learning." *Construction and Building Materials* 329 (2022): 127082.
- [3] Emad, Wael, et al. "Prediction of concrete materials compressive strength using surrogate models." *Structures*. Vol. 46. Elsevier, 2022.
- [4] Liu, Kexin, et al. "Development of compressive strength prediction platform for concrete materials based on machine learning techniques." *Journal of Building Engineering* 80 (2023): 107977.
- [5] Chi, Lin, et al. "Machine learning prediction of compressive strength of concrete with resistivity modification." *Materials Today Communications* 36 (2023): 106470.
- [6] Ghunimat, Dalin, et al. "Prediction of concrete compressive strength with GGBFS and fly ash using multilayer perceptron algorithm, random forest regression and k-nearest neighbor regression." *Asian Journal of Civil Engineering* 24.1 (2023): 169-177.
- [7] Liu, Gaoyang, and Bochao Sun. "Concrete compressive strength prediction using an explainable boosting machine model." *Case Studies in Construction Materials* 18 (2023): e01845.
- [8] Liu, Kexin, et al. "Development of compressive strength prediction platform for concrete materials based on machine learning techniques." *Journal of Building Engineering* 80 (2023): 107977.
- [9] Li, Hong, et al. "Compressive strength prediction of basalt fiber reinforced concrete via random forest algorithm." *Materials Today Communications* 30 (2022): 103117.
- [10] Imran, Hamza, et al. "Latest concrete materials dataset and ensemble prediction model for concrete compressive strength containing RCA and GGBFS materials." *Construction and Building Materials* 325 (2022): 126525.