FACTORS’ SELECTION EFFECT AND COMpressive
STRENGTH PREDICTION OF SCC USING A HYBRID NETWORK
BASED ON GA

Wei Liang1,2,3, Ming Lin1,3, Jiangfeng Dong2, Shucheng Yuan1,3

1. Sichuan Agricultural University, College of Civil Engineering, Dujiangyan, 611830, Chengdu, China.
2. Failure Mechanics & Engineering Disaster Prevention and Mitigation, Key Laboratory of Sichuan Province, Sichuan University, 610065, Chengdu, China.
3. Sichuan Higher Education Engineering Research Center for Disaster Prevention and Mitigation of Village Construction, Sichuan Agricultural University, Dujiangyan, 611830, Chengdu, China. E-mail addresses: liangwei9090950@163.com

ABSTRACT

Compressive strength is the most important evaluation index for concrete. In order to predict the compressive strength of self-compacting concrete, two kinds of artificial neural networks (ANNs), including the BP (Back-propagation) networks and the hybrid networks DRGA-BP (Dimension reduction back-propagation based on genetic algorithm), were designed and applied in this study. With DRGA-BP, the most representative variables were selected out from many initial inputs to reduce data dimensions and also the weights and thresholds of BP model were optimized. The results showed that the hybrid model presented better prediction accuracy with the R² (coefficient of determination) of 0.9602, and appeared to well agree with the experimental data and was quite reliable. Finally, a mix ratio design method based on DRGA-BP model was proposed for reducing material waste and saving time in the process of concrete production with continuous adjustment.

KEYWORDS

Self-compacting concrete, Artificial neural network, Genetic algorithm, Compressive strength

INTRODUCTION

Self-compacting concrete (SCC) is a kind of high-performance concrete that can flow through gaps of steel bars and fill corners of molds by its own weight, and can be compacted without additional mechanical vibration during the casting process.[1-2]. SCC has the advantages of eliminating noise of vibration, reduction in labor cost and increasing casting speed. For these reasons, SCC has been widely applied in concrete structures [3-5].

However, the performance of SCC is affected by many complex factors, and mixture ratio as one of the most important factors is still designed by empirical method [6]. At present, the mix design of SCC mainly includes the following steps: (1) Select coarse aggregate and its dosage; (2) Calculate the mortar dosage; (3) Calculate the amount of sand and the paste content; (4) Calculate the water-binder ratio according to the design strength of SCC; (5) Calculate the dosage of binding material according to the volume of paste, apparent density of cementing material, water-binder ratio, etc.; (6) Calculate the amount of water; (7) Determine the dosage of additives; (8) Trial matching and adjustment of mix ratio. Okamura et al.[1] suggested the following parameters: the
amount of coarse aggregate, which accounts for about 50% of the total aggregate volume; the amount of fine aggregate, which accounts for about 40% of the total aggregate volume; low water-cement ratio and use of superplasticizer. Domone[6] provided the median value of the mix ratio through a large number of data statistics: the amount of coarse aggregate, which accounts for 31.2% of the total volume; the amount of fine aggregate, which accounts for 47.5% of the total volume; the content of paste, which accounts for 34.8% of the volume; the dosage of binding material is 500kg/m³; the water-cement mass ratio is 0.34. The experience statistical data and empirical formulas, can offer some reference value for design of SCC mix proportion, and help to narrow the scope of each parameter. However, due to the variety of raw materials that can be used in SCC, especially the increasing use of mineral admixtures such as fly ash, silica powder and blast furnace slag, the influencing factors are more complex compared to common concrete. Therefore, various mixing ratio parameters still need to be adjusted and tested in practical engineering. In addition, there is a nonlinear relationship between these parameters and concrete strength, so the mix proportion design method based on empirical linear formula is no longer applicable. Therefore, if a more efficient and accurate mix ratio design method of SCC can be proposed, a large number of trial materials and time can be effectively saved.

Artificial neural network (ANN) is a model that imitate human brain to process data. Outside information is transmitted between neurons in ANN and adjusted by learning. Eventually, the trained ANN has an ability to react to the input data and obtain the desired output [7-8]. In recent years, ANN has been applied to many different fields, such as chemical engineering, medical industry, food technology and civil engineering, to solve nonlinear problems for modelling and optimization [9]. This method has also been applied to concrete strength prediction and mix ratio design in the past years. Lai et al.[10] concluded that the neural network performance was independent of the number of neurons in the hidden layer in the range of 4-8. Yeh [11-12] proposed a method of optimizing concrete mixture using ANN. For the prediction of concrete compressive strength, the Back-propagation (BP) neural network is the most used kind of ANN models [13-14]. Past research results show that the prediction results of ANN are more accurate than the traditional regression results[15], and the use of ANN makes it more convenient and simpler to analyze the influence of various factors on mix ratio. However, there are too many initial parameters that can be applied as input data in ANNs. These parameters are usually raw material types and their proportions, and most of them are chosen based on experience. Duan et al.[16] used 14 parameters as input data, such as water, cement, sand, natural aggregate, coarse aggregate, fineness modulus of sand, maximum size of coarse aggregate, water–cement ratio, type of coarse aggregate, impurity content, water absorption of coarse aggregate, saturated surface dry specific gravity of coarse aggregate, replacement ratio by volume and conversion coefficient. Ghafari et al.[17] used cement, sand, silica fume, quartz flour, water, superplasticizer and steel fiber as input data. Chithra et al.[18] applied six mix constituents as input data, such as cement content, nano silica content, fine aggregate content, copper slag content, age of specimen and super plasticizer dosage. These variable parameters affected the prediction accuracy a lot, but there were few investigations tried to find out the critical ones in the past. Due to the fact that the effect factors are many and not independent mutually, it is easy to occur the overfitting phenomena and cause the problems of low accuracy and time-consuming process of networks. In order to solve this problem, these factors need to be optimized and selected to find out the factors which can best reflect the relationship between inputs and outputs.

Genetic algorithm (GA) simulate the evolutionary processes of species in the nature. It can be used to acquire a global optimal solution to solve nonlinear problems and optimize the initial weights and thresholds in BP network [19]. In this paper, a hybrid ANN model, DRGA-BP (Dimension Reduction Back-propagation based on Genetic Algorithm), is constructed to optimize the inputs and improve the prediction accuracy of compressive strength of SCC.
DATA COLLECTION AND NORMALIZATION

The experimental data sets were obtained from the research of Nepomuceno [20], including 60 mixes and corresponding compressive strength data of SCC at 28 days. In this paper, 40 sets of data were used in the training set, while the other 20 sets of data were used to test the accuracy of the prediction.

The input data were divided into 10 categories, including the amount of raw materials (cement, mineral admixtures, water, superplasticizer, fine aggregate and coarse aggregate) and binding materials (cement and mineral admixtures), and the ratio among the raw materials (water-binder ratio, sand ratio and mineral admixtures replacement ratio). In addition, the type of cement was Portland cement with a density of 3040~3140 kg/m$^3$; the mineral admixtures included single or combined limestone powder, fly ash and granite ash, with densities of 2380~2720 kg/m$^3$; the superplasticizer was polycarboxylic acid superplasticizer with a density of 1050 kg/m$^3$; the fine aggregate was natural sand, with a density of 1598 kg/m$^3$ and a fineness of 2.22; the coarse aggregate was granite crushed stone, with a density of 1642 kg/m$^3$ and a fineness of 5.78. Table 1 shows the range of these input and output data sets.

<table>
<thead>
<tr>
<th>Parameters type</th>
<th>Code</th>
<th>Description</th>
<th>Dosages in kg for solids and in liters for water and SP per cubic meter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td></td>
<td></td>
<td>Minimum</td>
</tr>
<tr>
<td>C</td>
<td>Cement</td>
<td></td>
<td>276</td>
</tr>
<tr>
<td>M</td>
<td>Mineral admixture</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>W</td>
<td>Water</td>
<td></td>
<td>154</td>
</tr>
<tr>
<td>P</td>
<td>Superplasticizer</td>
<td></td>
<td>4.8</td>
</tr>
<tr>
<td>B</td>
<td>Binding material</td>
<td></td>
<td>547</td>
</tr>
<tr>
<td>S</td>
<td>Superplasticizer</td>
<td></td>
<td>656</td>
</tr>
<tr>
<td>G</td>
<td>Coarse aggregate</td>
<td></td>
<td>729</td>
</tr>
<tr>
<td>W/B</td>
<td>Water-binder ratio</td>
<td></td>
<td>0.24</td>
</tr>
<tr>
<td>SR</td>
<td>Sand ratio</td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td>RR</td>
<td>Mineral admixture replacement ratio</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

Output $f_{c,28}$ Compressive strength (MPa) 38.3 86.8 65.2

Because the units of the input data are different and some of the data value range is larger, if the collected data is directly inputted into the network, it will lead to the problem of long training time and slow convergence, and the range of activation functions is also limited. Therefore, it is necessary to normalize the data before training the network and map the data to the interval of [0,1] or [-1,1]. The normalization algorithm adopted in this paper was $y = 2 \times (x - \min) / (\max - \min) - 1$. ‘x’ is the input vector, ‘min’ is the minimum value of ‘x’, ‘max’ is the maximum of ‘x’, and ‘y’ is the normalized output vector.
IMPROVED ANN MODEL BASED ON GA

Genetic Algorithm

In this paper, GA and BP network are combined to improve the shortcomings of traditional BP network. At the same time, it is expected to screen out the factors that have little influence on the strength of self-compacting concrete by reducing dimension of independent variables, so as to simplify input data and improve learning efficiency and prediction accuracy.

The implementation process of genetic algorithm is shown in Figure 1. After selection, crossover and mutation, population P reaches the requirements of environmental fitness and forms a new population Pn. The three most important steps are as follows[21]:

1. Selection. Select individuals with high fitness from the current population. And these individuals will have a greater chance of being passed on to the next generation.
2. Crossover. Individuals in the group are randomly paired with each other to obtain new individuals.
3. Mutation. There is a certain probability that an individual's genes will change and a new individual will be born.

In this paper, the roulette algorithm is used to select individuals. The higher the individual fitness, the greater the probability of selection. Specific operations are as follows:

1. Calculate the probability of each individual being selected:
   \[
P(X_i) = \frac{f(X_i)}{\sum_{i=1}^{m} f(X_i)} \quad (1)
   \]

   where, \( f(X_i) \) is the fitness of individual \( i \); \( m \) is the population size, that is, the number of individuals in the population.

2. Calculate the cumulative probability of each individual:
   \[
   q_i = \sum_{i=1}^{m} P(X_i) \quad (2)
   \]

3. Generate the random number \( r \) in (0,1). If \( q_{i-1} < r \leq q_i \), then the individual \( i \) is selected.
4. Repeat the operation (3) for \( m \) times.

According to the above operations, if an individual has high fitness, then that individual will be selected many times, thus its gene ratio will expand in the population.
In crossover operations, arithmetic crossover is used. Pairs of individuals in the population are randomly paired, and for each pair of individuals, two new individuals are generated according to the following formula:

\[ c_1 = p_1 \times a + p_2 \times (1 - a) \]
\[ c_2 = p_1 \times (1 - a) + p_2 \times a \]

where, \( p_1 \) and \( p_2 \) are the parent individuals respectively; \( c_1 \) and \( c_2 \) are the offspring after the crossover operation; \( a \) is a random number between 0 and 1, namely the crossover probability.

In this paper, the nonuniform variation method is used. The process by which a gene in an individual is randomly disturbed in a small probability to produce a new individual. Suppose \( x = (x_1, x_2, \ldots, x_i, \ldots, x_n) \) is a set of chromosomes, that is, an individual, and one of the genes \( x_i \) is selected as the mutation object, so the new individual is \( x = (x_1, x_2, \ldots, x_i, \ldots, x_n) \). This process can be expressed as:

\[
x_i = \begin{cases} 
  x_i + \Delta(g, u_i - x_i) & \text{if } t = 0 \\
  x_i + \Delta(g, x_i - l_i) & \text{if } t = 1 
\end{cases}
\]

where, \( u_i \) and \( l_i \) are the upper and lower bounds of value respectively; \( t \) is a random value, 0 or 1. And \( \Delta(g, y) \) is calculated by the following formula:
\[
\Delta(g, y) = y \left( 1 - r \left( \frac{(1 - g)}{G} \right) \right)
\]  

(6)

where, \( g \) is the current number of evolutionary generations; \( G \) is the maximum number of evolutionary generations; \( b \) is the a parameter that determines the degree of nonuniform variation; \( r \) is a random number in the range \([0,1]\).

**DRGA-BP model**

The hybrid neural network DRGA-BP for dimension reduction is shown in Figure 2. After normalization processing, each group of data is input into the computing network as an individual. During population initialization, genes in each individual chromosome are randomly coded with '1' or '0'. In BPNN-A, which represents the back-propagation neural network “A”, the optimal weights and thresholds are obtained by genetic algorithm, and then substituted into BPNN-A as initial weights and thresholds for training. In process (a), the influence of initial weights and thresholds on the prediction results is reduced because the initial weights and thresholds are not obtained randomly. At the same time, the problem of BP neural network falling into local optimal is avoided. In process (b), several factors that have a greater impact on concrete strength are selected and input into BPNN-B (back-propagation neural network “B”) as input data. Finally, the data are trained in BPNN-B and the prediction results can be obtained.

In this paper, the reciprocal of the Sum of the Squared Errors (SSE) is taken as the fitness function:

\[
f(X) = \frac{1}{sse(T_p - T_a)} = \frac{1}{\sum_{i=1}^{n}(t_{p,i} - t_{a,i})^2}
\]

(7)

where, \( T_p = \{t_{p,1}, t_{p,2}, ..., t_{p,n_p}\} \) is the predicted value in the training phase; \( T_a = \{t_{a,1}, t_{a,2}, ..., t_{a,n_a}\} \) is the actual value in the training phase; \( n_i \) is the number of samples in the training phase.
After several iterations, when a certain termination condition is met, individuals with optimal fitness survive and form a new population. This means that the most representative and effective variables are selected from the initial input data. Then, the selected factors are used as input data for training and testing in the new BP network (BPNN-B). Finally, the prediction results are obtained.

**Final selected factors**

In this hybrid neural network, the BPNN was set as a three-layer neural network with 8 hidden neurons. Levenberg-marquardt algorithm was used to train the data. The training termination conditions of this hybrid neural network were set as follows:
(1) Maximum number of iterations: 1000
(2) Training target error: 0.001
(3) Maximum number of verification checks: 10

If any of the above conditions were satisfied, the training was terminated. The population size $m$ in Formula (1) was set as 8. To make sure that the whole process does not go on endless, the maximum evolutionary algebra was set as 50, that means every 8 individuals formed a population, and the evolution process stopped after up to 50 times even if the above termination conditions had not been met. The 10 factors listed in Table 1 were used as input data for prediction.

In process (b), typical evolutionary process is shown in Figure 3. As can be seen that the average fitness of individuals in the population increased with the growth of evolutionary generations, and the difference between optimal fitness and average fitness became smaller and smaller. Eventually, after many iterations, fitness values tended to be constant. It can be considered that the evolutionary process had been completed. And the individuals with the best fitness, that was, the factors that had the greatest influence on the compressive strength, formed new populations as optimized parameters.

![Typical evolutionary process](image)

**Fig. 3 – Typical evolutionary process**

**MODEL VERIFICATION AND COMPARATIVE ANALYSIS**

In BPNN-B, the best effect achieved in the training process was in the 14th training session, which mean-square error (MSE) was only 0.0009841. The training results are shown in Figure 4. As can be seen, the training results were very consistent with the actual values, which meant the training effect was good. Besides, in total of five variables as the final selected factors had survived in the evolutionary process, including cement (C), mineral admixture (M), water-binder ratio (W/B), sand ratio (SR) and mineral admixture replacement ratio (RR). It shows that these five factors have great influence on the compressive strength of SCC, and similar conclusion can be obtained from relevant literatures[22]. These five optimal variables, as input data, participated in network BPNN-B for compressive strength prediction.

In order to compare and confirm the five selected factors were reasonable and could improve the accuracy of the prediction, the five selected factors and all ten original factors respectively as two groups of input data were applied to a conventional three-layer BP neural network with eight of hidden neurons for test, namely BPNN-5 and BPNN-10, as listed in Table 2.
Fig. 4 – Training results of BPNN-B in the hybrid neural network DRGA-BP

The prediction results of the compressive strength with hybrid network DRGA-BP and comparison BP network BPNN-5 and BPNN-10 are shown in Table 2 and Figure 5. In the table, RMSE is the root mean square error, which is the square root of the squared deviation between the predicted value and the actual value and the ratio of prediction times 'n'. It is very sensitive to the very large or very small errors in a set of measurements, and can well reflect the precision of the prediction. The expression of RMSE is as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (T_i - \hat{P}_i)^2}
\]

where, \(P_i\) is the predicted value; \(T_i\) is the actual value; \(\overline{T}_i\) is the average of all actual values; \(n\) is the number of samples.

\[R^2 = 1 - \frac{\sum_{i=1}^{n} (T_i - \hat{P}_i)^2}{\sum_{i=1}^{n} (T_i - \overline{T}_i)^2}\]

where, \(R^2\) is the determining coefficient, which measures the overall fitting degree of the regression equation and represents the overall relationship between the dependent variables and all independent variables. The closer \(R^2\) is to 1, the higher the reference value of the relevant equations will be. The closer to 0, the lower the reference value. Its expression is:

RMSE is the root mean square error, which is the square root of the squared deviation between the predicted value and the actual value and the ratio of prediction times 'n'. It is very sensitive to the very large or very small errors in a set of measurements, and can well reflect the precision of the prediction. The expression of RMSE is as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (T_i - \hat{P}_i)^2}
\]

where, \(P_i\) is the predicted value; \(T_i\) is the actual value; \(\overline{T}_i\) is the average of all actual values; \(n\) is the number of samples.

It can be seen that the prediction results of hybrid network DRGA-BP proposed in this paper were closest to the real value and had the best prediction effect. When the same 10 factors were used as the initial input data, compared with network BPNN-10, the prediction results of DRGA-BP had a very high accuracy, and the RMSE was 2.2329 MPa and \(R^2\) was 0.9602. Besides, by comparing the prediction results of BPNN-10 and BPNN-5, it can be seen that using the selected factors as the initial input data in the same BP neural network had effectively improved the accuracy of prediction. In addition, the comparison test showed that the prediction effect of BP neural network was very sensitive to the initial input data, so the improved network DRGA-BP using genetic algorithm can effectively suppress the impact of initial weight and threshold on the prediction effect.
Tab. 2: Prediction results of ANN

<table>
<thead>
<tr>
<th>Code</th>
<th>Number of hidden neurons</th>
<th>Initial input factors</th>
<th>Optimized output factor as selected factors</th>
<th>RMSE (MPa)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPNN-10</td>
<td>8</td>
<td>C, M, W, SP, B, S, G, W/B, SR, RR</td>
<td>/</td>
<td>6.5197</td>
<td>0.6610</td>
</tr>
<tr>
<td>BPNN-5</td>
<td>8</td>
<td>C, M, W/B, SR, RR</td>
<td>/</td>
<td>4.2972</td>
<td>0.8527</td>
</tr>
</tbody>
</table>

Fig. 5 – The actual value versus the predicted value

MIX RATIO DESIGN METHOD BASED ON DRGA-BP MODEL

Five key factors, including cement (C), mineral admixture (M), water-binder ratio (W/B), sand ratio (SR) and mineral admixture replacement ratio (RR), were selected out by DRGA-BP as control factors affecting the compressive strength of SCC. The predicted values with DRGA-BP and experimental values (actual values with experiments) were respectively interpolated to draw the relationship diagram of water-binder ratio and replace ratio, water-binder ratio and sand ratio on the compressive strength of SCC, as shown in Figure 6 and Figure 7. It can be seen that the experimental values and the predicted values showed the same influence trend, and the value of water-binder ratio, replacement ratio and sand ratio was not a simple linear relation to the compressive strength of SCC. In addition, when the mineral admixture replacement ratio was higher than 0.48, the corresponding strength of SCC decreased significantly.
Fig. 7 – Effect of W/B and SR on the compressive strength of SCC

The influence relation of water-binder ratio, replacement ratio and sand ratio on the compressive strength can provide reference for the mix ratio design of SCC with a certain strength. The main steps are as follows: (1) Determine the design compressive strength of SCC; (2) Determine the total binder amount and aggregate amount; (3) Select water-binder ratio, sand ratio and mineral admixture replacement ratio; (4) Calculate the specific dosage of water, cement, mineral admixture, fine aggregate and coarse aggregate; (5) Determine the dosage of additives; (6) Trial matching and adjustment of mix ratio.

Taking the mix ratio design of SCC with 40 MPa of compressive strength at 28 day as an example, the total colloidal quantity is 560 kg/m³, the total coarse and fine aggregate quantity is 1600 kg/m³. And the control factors can be determined in a small range as follows: water-binder ratio is 0.31-0.32, sand ratio is 0.44-0.50, and mineral admixture replacement ratio is 40%-50%. Finally, the water-binder ratio is determined as 0.31, the sand ratio is determined as 0.50, and the mineral admixture replacement ratio is determined as 40%. The specific amount of each material is calculated: cement is 336 kg/m³, fly ash is 224 kg/m³, fine aggregate is 800 kg/m³, coarse aggregate is 800 kg/m³, and water is 174 kg/m³. By substituting the above data into the artificial neural network DRGA-BP, the 28-day strength is predicted to be 44.7 MPa, which can provide reference value for the preparation of SCC and effectively reduce the matching work and adjustment time.

CONCLUSIONS

In this paper, a new hybrid network DRGA-BP was studied to predict the compressive strength of SCC. The following conclusions could be drawn from this study:

1. DRGA-BP network improved by genetic algorithm can reduce the influence of initial input data, initial weight and threshold on the prediction results, and can predict the compressive strength of SCC with high level of accuracy. The prediction is quite reliable for application in concrete industry to decrease the time-consuming laboratory tests.

2. DRGA-BP network can effectively screen out the main factors that affect the strength of SCC, including cement, mineral admixture, water-binder ratio, sand ratio and mineral admixture replacement ratio.

3. DRGA-BP can provide reference for the design of SCC mix ratio. However, it still needs a lot of experimental data to study and verify, so as to improve its prediction accuracy continuously.
CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

[9] Astry, G.; Gullón, B.; Labidi, J.; Gullón, P. Comparison between developed models using response surface methodology (RSM) and artificial neural networks (ANNs) with the purpose to optimize oligosaccharide mixtures production from sugar beet pulp. IND CROP PROD 2016, 92, 290-299.