Strength Prediction Model of High Strength Fly Ash Concrete Based On Mixed Kernel Function

Zhongling Zhao\textsuperscript{1, a}, Jing Li\textsuperscript{2, b}, Xiwen Feng\textsuperscript{2, *}, Xiao Wang\textsuperscript{2, c}, Mengying Gao\textsuperscript{2, d}

\textsuperscript{1}Department of Finance and Economics, Shandong University of Science and Technology, Jinan250000, China
\textsuperscript{2}School of Mining and Safety Engineering, Shandong University of Science and Technology, Qingdao266000, China
*Corresponding author e-mail:644875535@qq.com, \textsuperscript{a}zzl6060@126.com, \textsuperscript{b}lijing6318@163.com, \textsuperscript{c}wangxiao_sdkd@163.com, \textsuperscript{d}1454725269@qq.com

Abstract. In this paper, the proportion of fly ash in cementitious material, the amount of cementitious material and the ratio of water to cement are selected, and the rough set theory is used to process the data. A mixed kernel function support vector machine prediction model with conditional attribute as input and decision attribute as output is established to explore the influence of three factors on the strength of high strength fly ash concrete. The prediction model with high prediction accuracy is selected for optimization. According to the existing test data, the variation law of concrete strength with proportioning parameters is summarized.

1. Introduction
At present, the scale of major foundation projects in China is unprecedented, which is bound to have higher requirements for concrete, the traditional concrete has been unable to meet the demand, thus put forward high-strength concrete. The strength of ordinary concrete can be predicted by a linear function with cement-water ratio as a single factor (that is, Boromi formula). However, for new types of concrete, such as fly ash concrete and high performance concrete, because the influencing factors are more complex, and there are even the interaction of many factors, the linear function is no longer applicable, and often shows a specific nonlinear law [1]. Hu Mingyu [2] adopts the model and learning algorithm of BP network and RBF network, and then applies it to the strength prediction and optimization design of high strength fly ash concrete. Wu Dehui [3] analyzed the influence of water-binder ratio, cement dosage, fly ash substitution rate and sand ratio on the strength of fly ash concrete by LS-SVM model. Li Qiling [4] establishes a neural network model between the strength of fly ash high strength concrete and three influencing factors, and then uses genetic algorithm to search for the highest compressive strength and the corresponding formula.

On this basis, this paper mainly studies the compressive strength of high strength fly ash concrete for 28 days, puts forward a concrete strength prediction model with mixed kernel function, trains and predicts the sample data according to the model, and optimizes the predicted data. The variation law between strength and ratio of concrete is summarized.
2. Related theory

2.1. Rough set theory
Rough set theory is a data analysis tool put forward by Polish mathematician Pawlak Z in 1982. It is mainly used to deal with ambiguity and uncertainty, and plays an important role in data mining [5-7]. Attribute reduction is one of the core contents of RST, and reduction and kernel are important parts of RST.

2.2. Support Vector Machine algorithm
The support vector machine algorithm maximizes the interval between the support vector sample points and the hyperplane by constructing the optimal hyperplane of the sample, and then obtains the decision model of the support vector machine [8]. According to the Kuhn-Tucker condition, finding a set of constraint parameters $\omega$ and $b$ maximizes the interval $\gamma$ from the two heterogeneous support vectors to the classification hyperplane.

$$\max_{\omega,b} \frac{2}{||\omega||}$$

s.t. $y_i(\omega^T x_i + b) \geq 1, \ i = 1, 2, \ldots, m$.

The Lagrangian function of the optimization problem is obtained by introducing the Lagrangian multiplier $\alpha_i \geq 0$.

$$L = \frac{1}{2}||\omega||^2 + \sum_{i=1}^{m} \alpha_i (1 - y_i(\omega^T x_i + b))$$

In the process of constructing the prediction model, the sample data should be normalized according to formula (3):

$$x'_j = \frac{x_j - x_{\min}}{x_{\max} - x_{\min}}$$

Formula: $x'_j$ - normalized sample value; $x_j$ - training or test set input value; $x_{\min}$ - training or test set minimum value; $x_{\max}$ - training or test set maximum value.

2.3. Decision Tree
Decision tree (DT) is an algorithm commonly used to analyze data, which is easy to operate and easy to use [9]. The algorithm is based on Occam's razor theory and expounds the concept of information entropy.

$$\text{Entropy}(S) = - \sum_{i=1}^{m} p_i \log_2 p_i$$

Where $S$ is the training set; $p_i$ ($i=1, 2, \ldots, m$) is the frequency at which category attribute $C$ with $m$ category labels appears in all samples.

2.4. Naive Bayesian classifier
The greatest advantage of naive Bayesian classifier is the high speed of receiving a large amount of data training and query [10]. It has a relatively simple explanation for the learning situation of the classifier, we can simply through the query learning to calculate some probability values to understand its classification principle. At the same time, the NBC model needs to estimate very few parameters, is not very sensitive to the missing data, and the algorithm is relatively simple.
In theory, NBC model has the smallest error rate compared with other classification methods.

3. Information processing

Table 1 selects 30 high strength fly ash concrete data points from reference [11] [12]. C1 as the proportion of fly ash in cementitious material, C2 as cementitious material dosage (kg/m$^3$), C3 as water-cement ratio, Y is listed as the compressive strength grade (MaP) of concrete measured in 28 days. On the basis of 56 groups of data known in reference 3 as the optimal ratio, in order to facilitate prediction, the selected data are divided into three grades, including high(H), low(L) and middle (M) grades. The corresponding relationship between ratio and compressive strength was recorded.

| Number | C1  | C2  | C3  | Y    |
|--------|-----|-----|-----|------|
| 1      | 0.4 | 490 | 0.34| L    |
| 2      | 0.4 | 600 | 0.28| M    |
| 3      | 0.143 | 525 | 0.361| M    |
| 4      | 0.143 | 577 | 0.329| M    |
| 5      | 0.09 | 550 | 0.282| M    |
| 6      | 0.182 | 682 | 0.278| H    |
| 7      | 0.5  | 522 | 0.32 | L    |
| 8      | 0.09 | 600 | 0.242| H    |
| 9      | 0.115 | 610 | 0.32 | M    |
| 10     | 0.167 | 600 | 0.262| M    |
| 11     | 0.1  | 583 | 0.3  | M    |
| 12     | 0.3  | 583 | 0.3  | L    |
| 13     | 0.4  | 533 | 0.31 | M    |
| 14     | 0.22 | 590 | 0.45 | L    |
| 15     | 0.143 | 459 | 0.301| M    |
| 16     | 0.107 | 630 | 0.28 | M    |
| 17     | 0.5  | 560 | 0.265| H    |
| 18     | 0.5  | 600 | 0.35 | L    |
| 19     | 0.17 | 522 | 0.237| H    |
| 20     | 0.085 | 590 | 0.35 | L    |
| 21     | 0.167 | 600 | 0.233| H    |
| 22     | 0.15 | 583 | 0.3  | L    |
| 23     | 0.4  | 563 | 0.32 | L    |
| 24     | 0.3  | 600 | 0.265| M    |
| 25     | 0    | 505 | 0.4  | M    |
| 26     | 0.088 | 570 | 0.32 | M    |
| 27     | 0.123 | 570 | 0.28 | M    |
| 28     | 0.09 | 600 | 0.265| H    |
| 29     | 0.42 | 540 | 0.33 | L    |
| 30     | 0.23 | 600 | 0.23 | H    |

Because the three influencing factors selected above have different units and sizes, and there are order of magnitude differences between the dimensions, resulting in large errors, it is necessary to normalize, discretize and reduce the original data.

Delete duplicate data 1,5,8. The importance of C1, C2 and C3 was 0.037, 0.074 and 0.037, respectively.
4. Construction of prediction model

4.1. Establishment of prediction model of machine learning algorithm

After reduction, the data is less and has the characteristics of low dimension, so the machine learning algorithm is introduced, the conditional attribute C is used as the input and the decision attribute Y is used as the output to establish the intensity grade prediction model of different machine learning algorithms.

Naive Bayesian classifier (NBC) [13] prediction model: establish prediction model predict (NBC.fit), C4.5 decision tree (DT) [14] prediction model: establish prediction model predict (DT.fit); Support vector machine (SVM) prediction model: the radial basis (Radial Basis Function, RBF) kernel function is used as the SVM kernel function to establish the prediction model (SVM.fit).

4.2. Result analysis

The first 19 groups after reduction processing are taken as the training set, and the remaining 8 groups are used as the test set. The prediction model of the machine learning algorithm is used to predict the test set. The prediction results are shown in Table 2.

| Num | Actual | NBC | DT | SVM |
|-----|--------|-----|----|-----|
| 1   | 1      | 1   | 2  | 1   |
| 2   | 2      | 2   | 2  | 2   |
| 3   | 2      | 2   | 2  | 2   |
| 4   | 2      | 2   | 2  | 2   |
| 5   | 2      | 2   | 2  | 2   |
| 6   | 3      | 2   | 2  | 2   |
| 7   | 1      | 1   | 2  | 1   |
| 8   | 3      | 3   | 2  | 3   |

As can be seen from Table 3, the prediction accuracy of NBC and SVM prediction model is 82.5%. In order to further select the appropriate prediction model, the node error rate is calculated as shown in Table 3.

|                  | NBC   | DT    | SVM   |
|------------------|-------|-------|-------|
| Mean absolute error | 15.58% | 22.31% | 12.06% |
| Root mean square error | 29.66% | 24.63% | 16.24% |
| Absolute value of relative error | 33.09% | 50.48% | 11.03% |
| Relative root mean square error | 50.52% | 61.67% | 33.32% |

By comparison, the node error rate of the SVM prediction model is lower than that of the other models mentioned above. Considering the accuracy and node error rate, the prediction model of SVM is better.

5. Model optimization

SVM prediction model has strong learning ability, but the generalization ability is weak, and the prediction accuracy needs to be further improved. The RBF kernel function with strong learning ability is selected to establish the prediction model.
ability and the sigmoid kernel function with strong generalization ability are combined into mixed kernel function.

\[ K_{m}(x,x_i) = \lambda_1 \tanh(\beta x^T x_i) + \lambda_2 \exp\left(-\gamma \|x-x_i\|^2\right) \] (5)

In formula (5), \( \lambda_1 \) is the coefficient of Sigmoid kernel function, \( \lambda_2 \) is the coefficient of RBF kernel function, \( \lambda_1, \lambda_2 > 0 \), and \( \lambda_1 + \lambda_2 = 1 \). The prediction model of mixed kernel function support vector machine for high strength fly ash concrete is established, the test set is predicted, and the prediction accuracy curve of mixed kernel function support vector machine prediction model with \( \lambda_1 \) is obtained, as shown in figure 1.

![Fig. 1 The accuracy with the change curve of \( \lambda_1 \)](image)

From figures 1 and 2, we can see that the mixed kernel function support vector machine prediction model has strong learning ability and generalization ability, and the prediction accuracy is 100% when \( \lambda_1 = 0.05 \) and \( \lambda_2 = 0.95 \).

6. Conclusion
In this paper, the proportion of fly ash in cementitious material, the amount of cementitious material and the ratio of water to cement are selected, and the prediction models of NBC, DT and SVM are established.

In this paper, a hybrid kernel function support vector machine prediction model with conditional attribute as input and decision attribute as output is proposed, and the prediction model is better than other prediction models, which has a certain reference value for the mix design of concrete in the future.

References
[1] He Shiqin, Wang Haichao. Orthogonal experimental study on mix design of high performance concrete [J]. Industrial Building, 2003, 33 (8): 79779.
[2] Hu Mingyu, Tang Mingshu. Application of Neural Network in strength Prediction and Optimization Design of High strength fly Ash concrete [J]. Concrete, 2001 (01): 13-17.
[3] Zhang Chunyu, Li Qiling. Optimal design of fly ash concrete strength [J]. Journal of Tongji University (Natural Science Edition), 2003 (04): 456459.
[4] Wu Dehui. Prediction model of fly ash concrete strength based on LS-SVM [J]. New Building Materials, 2007 (03): 6670.5
[5] Vidhya K.A., Geetha T.V. Rough set theory for document clustering: A review [J]. Journal of Intelligent & Fuzzy Systems, 2017, 32(03):2165-2185.
[6] Wang Guangyue, Cui Haili, Li Qian. Study on determination method of Factor weight in Slope Stability Evaluation based on rough set Theory [J]. Geotechnical Mechanics, 2009, 30 (08): 2418-2422.

[7] Yu Hong, Wang Guoyin, Yao Yiyu. Research status and Prospect of decision rough set Theory [J]. Journal of computer Science, 2015, 08: 1628-1600.

[8] Liang Liming, Zhong Zhen, Chen Zhaoyang. Research and Simulation of Kernel function selection of support Vector Machine [J]. Computer Engineering and Science, 2015, 37 (06): 1135-1141. [2017-08-04].

[9] Feng Shaorong. Research and improvement of decision tree algorithm [J]. Journal of Xiamen University (Natural Science Edition), 2007 (04): 496500.

[10] High rock. Research on the improvement of naive Bayesian Classifier [D]. Guangzhou: South China University of Technology, 2011.

[11] Su Zheng. Preparation and Application of 6C0 pumping concrete mixed with fly Ash [J]. Concrete, 1996, 2: 31-32.

[12] Ma Jianxin, etc. Preparation of plastic High fluidity High strength concrete with Water reducing fly Ash and Control of slump loss [J]. Concrete, 1997, 1: 14-19.

[13] Shi Yongkui; Li Pengrui; Wang Ying; et al. Classification and Prediction of Rock Burst Based on BP Neural Network[J]. Electronic Journal of Geotechnical Engineering, 2015, 20 (13): 5839-5848.

[14] Cui Wei, Guo JingFeng, Zhang WenYuan, et al. Kind of Effieient Relational labe Naive Bays C lassification Algorithm. Science Technology and Engineering, 2008; 8 (16): 4706—4708