Estimation Of Compressive Strength Of Concrete Containing Manufactured Sand By Random Forest

Hai-Bang Ly, Van Quan Tran

Abstract: The compressive strength of concrete is an important parameter that required a precise manner to forecast before any fabrication process to save time and cost. In this work, the possibility of using Random Forest, as a machine learning algorithm, to predict the compressive strength of manufactured sand concrete was investigated. A reliable dataset containing experimental results from the available literature was collected and used to construct and validate the RF black-box. The results of this study encouraged the use RF to predict the compressive strength of concrete containing manufactured sand, as the correlation between the actual and predicted data was 0.966. Moreover, the curing age and the water to binder ratio were found as the most affecting factors to the prediction process of RF. This study might provide a reliable and quick numerical tool for engineers as to avoid worthless experiments.

Keywords : manufactured sand, concrete, compressive strength, random forest, machine learning

1. INTRODUCTION

The use of manufactured sand in concrete, an alternative for natural sand, has gained increasing attention due to the deficiency of suitable rival sand in most part of the world [1]. In general, manufactured sand is generated from hard stone by crushing. Some unique features of manufactured sand could be advantageous for applications in civil engineering related fields, for instance, they exhibited a cubical shape with grounded edges. Moreover, the negative point while using natural sand lies in the alkali-silica reactions with the alkali hydroxides in cement components. This might cause severe problems to structural members containing concrete such as expansion or deterioration. With manufactured sand, the selection of prequalification rocks with low silica content might be performed before the crushing procedure, which preventing unfavorable reactions occurred. Last but not least, the use of manufactured sand could enhance and improve the concrete strength and durability, thanks to the greater particles interlocking degree. Due to these beneficial characteristics, the quantification of several mechanical behaviors of concretes containing manufactured sand is crucial, especially the compressive strength.

Indeed, the compressive strength, reflecting the resistance to failure under compressive loading, is an important parameter that determines the concrete performance during service conditions. In addition to researches on manufactured sand concrete workability and durability, the compressive behaviors are also the subject of widespread investigations. Based on digital images and laboratory tests, Shen et al. [2, 3] pointed out the influence of the roughness and form of manufactured sand on the compressive strength of concrete. They were identified as having less effect on the compressive strength than the stone powder. Besides, higher water reducing admixtures content was required in manufactured sand concrete compared with river sand concrete. Moreover, nearly all the experiments showed that manufactured sand concrete had higher strength than concrete containing river sand using similar paste composition. In many other studies such as the works of Park [4] or Donza et al. [5], identical observations were also mentioned. In general, experimental approach to investigate the mechanical properties of manufactured sand concrete required an important number of tests to be able to fully understand the effect of all the mixture components. So that it is highly priced and necessitated a long waiting period from the design, casting steps to the final mechanical tests. A powerful numerical tool based on machine learning algorithm could be an alternative manner to perform the simulations without conducting unnecessary experiments. Undeniably, the use of machine learning algorithms has been widely applied in civil engineering applications nowadays, providing promising opportunities to a sustainable development in the construction sector [6]. Various machine learning algorithms have been developed and applied to solve real world problems such as Adaptive Network based Fuzzy Inference System (ANFIS) based model to deal with pavement materials [7], road surface roughness [8], traffic air pollution [9], buckling of structural members [10] or using Artificial Neural Network for predicting concrete properties [11]. Overall, nonlinear and complex relationships between the mixture components versus the corresponding concrete compressive strength could be detected in a soft computing manner thanks to the strong ability of machine learning algorithms. While different algorithms were used to investigate the compressive strength of manufactured concrete in the literature, it is found that tree based learning algorithms have not been fully investigated to solve such interesting problem. Therefore, the main objective of this work is to study the possibility of applying Random Forests (RF) to predict the compressive strength of concrete using manufactured sand. A brief introduction of RF is presented in Section 2, following by the presentation of the concrete dataset in Section 3. The prediction results are given in Section 4. Finally, several conclusions and perspectives are addressed in Section 5.

2 METHOD USED

2.1 Random Forest

Random Forest (herein denoted as RF) is a machine learning algorithm usually used to solve classification and regression problems [12]. The RF algorithm was first introduced in 2001...
While comparing with the ability in predicting of a single decision tree, RF could reduce the prediction variation and errors, as well as improve the prediction performance. The nature of RF lies in the fact that RF combines many decision trees in a special manner, the so-called Bagging approach. The classical approach of RF provides an ability to learn in parallel from a multiple of decision trees that are randomly constructed and trained with many subsets containing different samples. Each tree in the forest is trained by a subset with randomly distributed data following the bagging principle and also with random features. The final results are given as the average value of each single decision tree for a regression problem, or defined by a majority vote for a classification problem. The RF algorithm is well known by the so-called out-of-bag, which is a method for quantifying the prediction error of the model. Due to many advantages of RF, this algorithm has been widely applied in various applications [13]. For the detailed development of RF, different steps of the algorithm, the readers could referred to the literature [12].

2.2 Performance criteria
In this study, classical statistical measurements of error were used, namely the root mean square error (RMSE) and the Pearson correlation coefficient (R). Generally, lower values of RMSE indicate the better accuracy of any machine learning algorithm. On the contrary, higher values of R represent better correlation between the predicted and actual data, thus better prediction accuracy. The range of value of R are from -1 to 1, whereas the minimum value of RMSE is 0. The formulations of these criteria could be referred to the literature [14].

3 DATA USED
The data used in this study were gathered from the available literature [15]–[18]. The target of the RF algorithm was the compressive strength (MPa) of manufactured sand concrete. It is worth noticed that in the given dataset, both cubic and cylinder compressive strength were obtained. A factor of 0.82 was applied to convert these values for further usage in the development of the RF model. The number of input used in this study is 11 variables. They were the compressive strength of cement (MPa), the tensile strength of cement (MPa), the curing age (day), the maximum diameter of the aggregate (mm), the content of stone powder (%), the fineness modulus of sand, the ratio of water to binder, the water to binder ratio, the content of water (kg/m³), the sand ratio (%) and the slump (mm). A number of 298 samples was finally collected. The data were then divided into the training dataset (containing 70% of the data) and the testing dataset (30% of the data) aiming at constructing the RF model as well as the validation phase once the model was built.

4 RESULTS AND DISCUSSION
4.1 Performance of RF
This section demonstrates the performance of RF algorithm in predicting the compressive strength of manufactured sand concrete. In RF algorithm, the number of trees is an important parameter that directly affect the prediction accuracy. An appropriate number of decision trees could ensure the accuracy in predicting the output of RF. Fig. 1 shows the out-of-bag error in function of the number of trees in RF. As can be seen, a number of 50 trees could guarantee the low error values and the rate of convergence decreased when increasing the number of trees. It could be concluded that, in this study, the use of 500 grown trees was sufficient to achieve reliable prediction results.

In the following parts, the prediction results of RF are presented. Fig. 2 shows the histogram of error between the actual values of compressive strength of manufactured sand concrete versus those obtained from experiments. As can be seen, a high concentration of error around 0 was obtained, showing the efficiency of RF. Several higher values of error were found (i.e. -20, -15) but only 2 cases were observed. This clearly showed that, for the training part, the RF algorithm performed well in solving the problem.

Next, the performance of the testing dataset is presented. Fig. 3 shows the histogram of error between the actual and predicted output. Again, only 4 values of error superior to 10 (Mpa) were observed. The remaining errors were highly concentrated between 0 and ±5 (MPa). The values of RMSE were 3.8699 and 4.9962 for the training and testing datasets, respectively. The values of mean error were computed as -0.12604 and -0.7388 for the training and testing parts, respectively. Idem, values of standard deviation (denoted as St.D.) were 3.8775 and 4.9699 for the training part and testing one, respectively. It can be seen that the accuracy (i.e. error mean, standard deviation error and RMSE) of the training data was superior to the testing one, thus helpful in preventing overfit problem. Moreover, a number of 500 decision trees was
selected as to avoid overfitting. Thus, the results were reliable and the RF model could be used for further investigation.

The regression analysis is next conducted to show the performance of RF in predicting the compressive strength of manufactured sand concrete. Fig. 4 shows the regression graph for the training dataset, whereas Fig. 5 displays that of the testing dataset. As can be seen, the RF algorithm showed excellent accuracy in terms of R, where values of R = 0.97736 and R = 0.966 were obtained for the training and testing datasets, respectively. Overall, the results indicated that RF was a good predictor for predicting the compressive strength of manufactured sand concrete. The use of RF to predict such mechanical property is completely feasible and could be a promising predictor to estimate other important mechanical properties of manufactured sand concrete.

4.2 Feature importance analysis

Feature importance analysis is performed in this section. Basically, while constructing and developing the RF model, this algorithm could provide information related to the sensitivity of the features used to predict the output. The importance of features is computed as the sum of changes in the risk made by the splitting process then divide by the sum of the number of branch. In general, the out-of-bag is the criterion to deduce the importance of feature. The higher the value of out-of-bag, the more important the input. Fig. 5 shows the sensitivity analysis of 11 input parameters in this study in predicting the compressive strength of manufactured sand concrete. It can be seen that the curing age and the water to binder ratio were the two most affecting features to the prediction of compressive strength. As the range of the curing age in this study was large (i.e. from 3 to 388 days), it was understandable to conclude that this input exhibited an important role in the compressive strength of manufactured concrete. Second, like any other type of concrete, the water to binder ratio is a crucial parameter that directly affects the mechanical properties of concrete. The water to cement ratio, water content and sand ratio were the three following important parameters to the prediction problem. Again, the water to cement ratio was an important parameter. In addition to the previous feature (i.e. water to binder ratio), it could be concluded that the contents of water and binder were very important. The amount of sand was at 5th position in order of sensitivity to the predicted output. It is interesting noticed that the main factors that differ manufactured sand and natural sand were not considered as input. However, the prediction accuracy was rather satisfactory. This showed that the use of manufactured sand to replace natural sand was possible and only the amount of sand in concrete was important. Besides, the slump, compressive strength of cement, finesse modulus of sand, tensile strength of cement, content of stone powder, maximum diameter of crushed stone were listed as unimportant variables. The out-of-bag errors of these variables were insignificant. Overall, the sensitivity analysis using RF feature importance analysis is helpful in finding several high impact parameters to the prediction process. Based on the results, further prediction of manufactured concrete compressive strength might be improved by selecting the appropriate features.
Fig. 6. Feature importance analysis of 11 input parameters in this study.

5 CONCLUSION

In this study, the possibility in using Random Forest algorithm to predict the compressive strength of manufactured sand concrete was investigated. In order to construct the database for the use of RF algorithm, reliable experimental results were gathered from the available literature. The dataset was next divided into the training (70% of data) and testing datasets (30% of the remaining data). Several well-known statistical criteria were used to evaluate the performance of RF algorithm in predicting the compressive strength of concrete containing manufactured sand. The results of this work showed that RF was a promising machine learning algorithm, as the correlation of the training and testing datasets were 0.97736 and 0.966, respectively. Besides, the obtained values of RMSE were 3.8699 and 4.9962 for the training and testing datasets, respectively. Finally, it was found that the curing age and the water to binder ratio were the two important parameters in predicting the compressive strength of manufactured sand concrete. Overall, the results in this study confirmed the ability of RF in predicting the concrete compressive strength, and maybe applicable to other related properties. The RF model is another alternative for the problem and can be used for further investigation of manufactured sand concrete.

REFERENCES


