ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

Prediction of Concrete Carbonation Depth using Decision Trees

Woubishet Zewdu Taffese, Esko Sistonen and Jari Puttonen

Aalto University, Department of Civil and Structural Engineering
P. O. Box 12100, FI-00076 Aalto, Finland

Abstract. In this work, three carbonation depth predicting models using decision tree approach are developed. Carbonation, in urban areas is often a reason for reinforcement steel corrosion that causes premature degradation, loss of serviceability and safety of reinforced concrete structures. The adopted decision trees are regression tree, bagged ensemble and reduced bagged ensemble regression tree. The evaluation of the predictions performance of the developed models reveals that all the three models perform reasonably well. Among them, reduced bagged ensemble regression tree showed the highest prediction and generalization capability.

1 Introduction

Corrosion of reinforcement steel in concrete induced by carbonation is the foremost cause of premature degradation, loss of serviceability and safety of reinforced concrete structures [1, 2]. Carbonation of concrete is a natural physicochemical process caused by the penetration of carbon dioxide from the surrounding environment into the concrete through pores in the matrix where the carbon dioxide reacts with hydrated cement. Calcium hydroxide ($Ca(OH)_2$) in contact with carbon dioxide (CO_2) forms calcium carbonate ($CaCO_3$). This chemical reaction reduces the alkalinity of the pore fluid from pH value around 13 to pH value of below 9. Consequently, the passive oxide layer steel reinforcement is destroyed and eventually corrosion of the steel bars will be initiated [2, 3].

Concrete carbonation depth at a given time in steady state conditions can reasonably be estimated using Eq. (1) for usual life-time of concrete structures. This equation is based on Fick's second law of diffusion and it is well known [2].

$$x = C\sqrt{t} \qquad (1)$$

where, x is the depth of carbonation at time t [mm] , C is coefficient of carbonation [$mm/d^{0.5}$], and t is the duration of carbonation [d].

Coefficient of carbonation is a decisive factor in determining carbonation depth. It is analyzed either by an accelerated carbonation test or by measuring the development of the carbonation depth from an existing concrete structure. Since carbonation is a slow process, it is usually investigated by performing accelerated test with a higher CO_2 concentration in a controlled environment at the age of 28 days [4]. Then, the measured carbonation depth is used to calculate the equivalent carbonation coefficient using Eq. (1). Carbonation coefficient is mainly controlled by diffusion of CO_2 into the concrete pore system. CO_2 diffusion through concrete depends on several factors such as CO_2 concentration, environmental condition, and concrete characteristics. Therefore, carbonation coefficient may significantly vary from one

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

concrete structure to another depending on environment and microstructural parameters which are linked with concrete composition and type of materials used.

Developing analytical carbonation depth prediction model is a challenging task since it is a function of many parameters that are complex to describe mathematically. Hence, building a model that can learn from readily available real data using a machine learning algorithms is a better alternative. Even though this approach is becoming a common practice in various engineering fields, its application in concrete durability is yet limited. Among several machine learning techniques, only artificial neural network is widely used in this research area, for instance, chloride penetration in concrete [5] and hygrothermal forecasting in thick-walled concrete [6].

This paper presents a machine learning method, namely a decision tree, for prediction of concrete carbonation depth.

2 Data understanding and preparation

2.1 Data understanding

Experimental data obtained from [7] is used to develop a model for predicting the depth of carbonation. This data were prepared for Finnish DuraInt-project. The project was carried out in cooperation between Aalto University and VTT Technical Research Centre of Finland. The data consists of concrete mixture ingredients and fresh and hardened properties of 46 specimens. Carbonation depths for half of the concrete specimen were conducted at the age of 28 days and the remaining half at the age of 56 days. The accelerated carbonation tests were performed by applying CO_2 of 1% in a controlled environment (temperature 21°C and relative humidity 60%) in accordance with EN 13295. The data contain both numerical and categorical inputs. In this work, only data of the concrete mixture ingredients and the carbonation depth is used, which is in total 15 features. These are: cement type, water to binding ratio (w/b), cement, blast-furnace slag (BFS), fly ash (FA), total effective water, total aggregate, aggregate < 0.125mm, aggregate < 0.25mm, aggregate < 4mm, product name of plasticizer, plasticizer, product name of air-entraining agent, air-entraining agent, carbonation period and carbonation depth.

2.2 Data preparation

An input matrix of [46x15] predictor values from concrete mixture parameters was arranged. Each column of an input matrix represents one variable, and each row represents one observation. A numeric column vector, carbonation depth, with the same number of rows as input matrix was prepared and assigned as a target. Each entry in output vector is the response to the data in the corresponding row of the input matrix. Since the environmental conditions for all test specimens were identical, this parameter is not included in the predictor matrix. The dataset were used for both training and testing datasets with 10-fold cross-validation.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

3 Modeling carbonation depth using decision trees

Decision tree is a nonparametric hierarchical data structure which implements the divide-and-conquer strategy. It is composed of internal decision nodes and terminal leaves as illustrated in Figure 1. The left panel plots the data points and partitions and the right panel shows the corresponding decision tree structure. Each decision node implements a test function with discrete outcomes labeling the branches. Given an input, at each node, a test is applied and one of the branches will be chosen depending on the outcome. This process starts at the root and is repeated recursively until a leaf node is hit, at which point the value written in the leaf constitutes the output [8].

In this work, three different decision trees are used to predict concrete carbonation depth. These are regression tree, ensemble bagged regression tree and bagged regression tree after features reduced. All the trees were developed using Matlab.

3.1 Regression tree

The structure of the regression tree is the same as that of the tree presented in Figure 1. The only difference is the leaves which contain real numbers instead of class labels. The regression tree is trained over the training dataset. The performance of the developed tree is measured by mean square error (MSE) and mean absolute error (MAE) on both training and testing dataset. MSE, the mean square error between predicted output ($\hat{Y}_i$) and target (Y_i), is the most common measure of accuracy, Eq. (2). The MAE of Eq. (3) is the more intuitive measure and is less sensitive to outliers.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \hat{Y}_i\right)^2 \qquad (2)$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|Y_i - \hat{Y}_i\right| \qquad (3)$$

where $\hat{Y}_i$ is the predicted output value, Y_i is the measured target value, and N is the number of observations.

The resulting MSE values for training and test dataset were 0.0416 and 4.3108, respectively. Significant difference in MAE of training and testing dataset is also observed. All these show that the developed regression tree generalized the test data poorly because it overfitted the training data as seen in the regression plot, Figure 2.

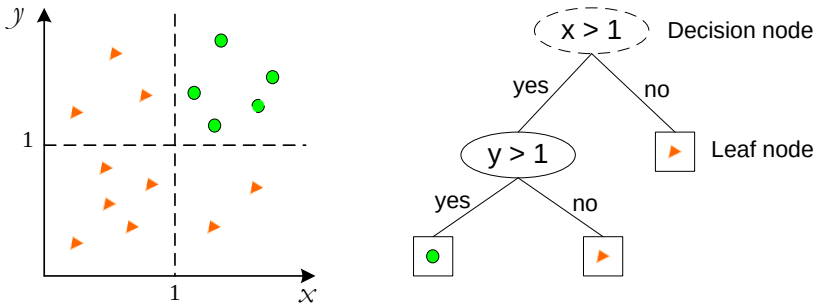

Fig. 1: Example of a dataset and the corresponding decision tree.

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

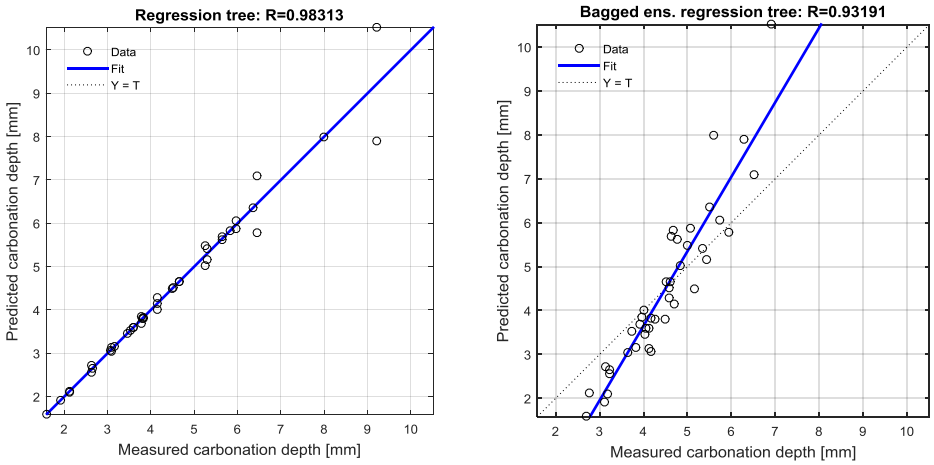

Fig. 2: Regression plot of predicted vs measured carbonation depth on training dataset for regression tree (left) and bagged ensemble regression tree (right).

3.2 **Bagged ensemble regression tree**

Bagging is one of the most effective methods that can be used to improve the predictive performance of a tree model by reducing the variance associated with prediction. This technique draws multiple bootstrap samples from the training dataset and generates multiple predictor trees, and then, the results are combined by averaging to obtain the overall prediction [9, 10].

An ensemble of bagged regression tree was developed with an initial default tree and leaf size. The performance evaluation indicates that ensemble of bagged regression tree has a high generalization capacity than the regression tree presented in Section 3.1. The MSE of the training and testing dataset was 0.9701 and 2.7223. Regression plot of predicted vs measured carbonation depth on training dataset for bagged ensemble regression tree is shown in Figure 2.

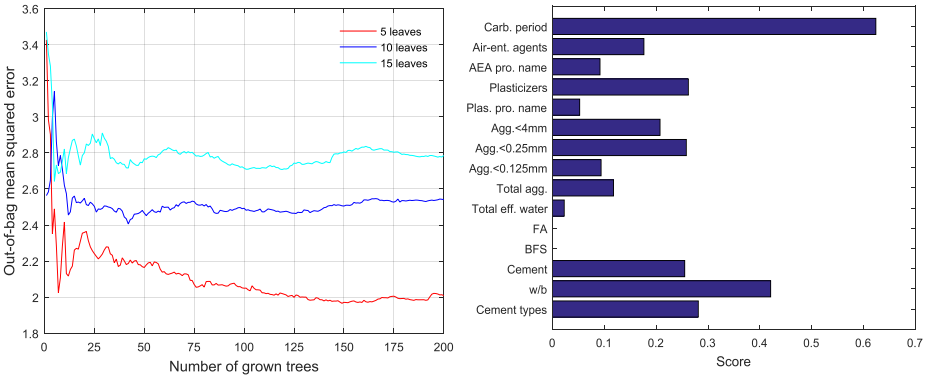

Fig. 3a: Out-of-bag mean square error vs number of grown trees (left). 3b: Relative importance of the input variables of the bagged ensemble regression tree (right).

ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 22-24 April 2015, i6doc.com publ., ISBN 978-287587014-8.
Available from http://www.i6doc.com/en/.

3.3 **Reduced bagged ensemble regression tree**

In order to minimize the prediction error of the bagged ensemble, we compute predictions for trees with different leaf sizes on its out-of-bag observations, Figure 3a. It can be observed that the out-of-bag error decreases well with the number of grown trees for leaf size of five. The relative importance of the input variables of the bagged ensemble regression tree is illustrated in Figure 3b. It can be clearly seen that the carbonation period and w/b are the foremost influential predictors for this dataset. Next to these variables, amount and types of cement, plasticizer and the distribution of aggregate play considerable role in predicting the carbonation depth for this dataset. This is a useful finding because plasticizer and aggregate distribution were overlooked in several existing analytical models.

After determining good predictors and an ensemble size from the out-of-bag error, a new bagged ensemble regression tree was constructed to enhance its performance further. In this case, the optimal number of leaf and trees was chosen as 5 and 150, respectively. Two parameters, BFS and FA, were reduced out of the total 15 features since they are unimportant to predict the carbonation depth in this dataset. The MSE of training and testing dataset of this model was 0.9536 and 2.2990. Figure 4 illustrate the predicted and the measured carbonation depth with the predicted error.

3.4 **Performance comparison**

An average of five round statistical performance measurements of all the carbonation prediction models are listed in Table 1. As shown in this table, reduced ensemble bagged regression tree is statistically outperformed all the other models for this dataset. The MAE values of this model for training and test dataset are 0.4755 and 0.5261, respectively. These indicate that this model reasonably fits the measured data and has relatively better generalization capability. All the performance measurements of the models are valid only for the considered specific dataset. If a different dataset is employed, the performance may differ noticeably. Generally, this study revealed the applicability of decision tree based models to predict concrete carbonation depth. As part of future work, the model will be evaluated using more experimental data.

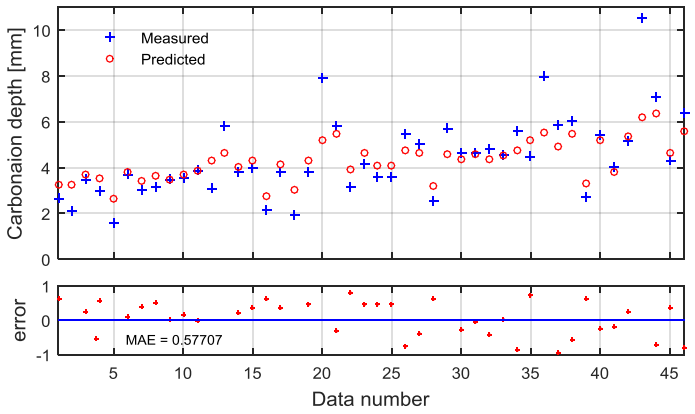

Fig. 4: Measured and predicted carbonation depth using bagged ensemble regression tree with the prediction error.

Models	MSE		MAE	
	Train	Test	Train	Test
Regression tree	0.0416	4.3108	0.0740	1.3437
Bagged ensemble regression tree	0.9701	2.7223	0.4927	0.6283
Reduced bagged ensemble regression tree	0.9536	2.2990	0.4755	0.5261

Table 1: Performance comparison of carbonation depth prediction models.

4 Conclusions

Three concrete carbonation depth prediction models based on decision tree method are presented. To develop the models, three different decision trees were adopted. They are regression tree, bagged ensemble regression tree and reduced bagged ensemble regression tree. The models prediction capacity was examined based on mean square errors and mean absolute error. Models developed using bagged ensemble with and without features extraction predict the carbonation depth with reasonably low error. The model developed using the former method has superior performance with relatively better generalization capability. This confirms the advantage of feature and ensemble size selection in improving performance. Furthermore, the bagged ensemble regression tree identified important variables that influenced the carbonation rate which was not considered in the existing analytical models. The models have potential to be part of a service life management system.

References

[1] K. Y. Ann, S.-W. Pack, J.-P. Hwang, H.-W. Song and S.-H. Kim, Service life prediction of a concrete bridge structure subjected to carbonation, *Construction and Building Materials*, 24:1494–1501, Elsevier, 2010.

[2] fib, Structural concrete: Textbook on behaviour, design and performance, 2nd edition, Volume 1, Technical Report, Fédération internationale du béton (*fib*), Lausanne, Switzerland, 2009.

[3] E. Sistonen, *Service life of hot-dip galvanised reinforcement bars in carbonated and chloride-contaminated concrete*, *Doctoral Thesis*, *Helsinki University of Technology, Espoo*, Finland, *2009.*

[4] P. Schiessl and S. Lay, Influence of concrete composition, In H. Böhni, editor, *Corrosion in reinforced concrete structures*, pages 91-134, Woodhead Publishing Ltd, Cambridge, 2005.

[5] H.-W. Song and S.-J. Kwon, Evaluation of chloride penetration in high performance concrete using neural network algorithm and micro pore structure, *Cement and Concrete Research*, 39:814–824, Elsevier, 2009.

[6] W. Z. Taffese, F. Al-Neshawy, J. Piironen, E. Sistonen and Jari Puttonen, Monitoring, evaluation and long-term forecasting of hygrothermal performance of thick-walled concrete structure, In *proceedings of OECD/NEA WGIAGE Workshop on the Non-Destructive Evaluation of Thick-walled Concrete Structures*, OECD, Pages 121-143, September 17-19, Prague, (Czech Republic), 2013.

[7] H. Kuosa, Concrete durability field testing: field and laboratory results 2007-2010 in DuraInt-project, Research Report, VTT Technical Research Centre of Finland, Espoo, Finland, September 2011.

[8] J. Gama, *Knowledge discovery from data streams*, Taylor and Francis Group, Boca Raton, 2010.

[9] E. Alpaydin, *Introduction to machine learning*, 2nd edition, The MIT Press, Cambridge, 2010.

[10] C. D. Sutton, Classification and regression trees, bagging, and boosting, In C. R. Rao, E. J. Wegman and J. L. Solka, editors, *Data mining and data visualization*, pages 303-330, Elsevier B.V., Amsterdam, 2005.