Construction and Building Materials 73 (2014) 771–780

Contents lists available at ScienceDirect

Construction and Building Materials

journal homepage: www.elsevier.com/locate/conbuildmat

Machine learning in concrete strength simulations: Multi-nation data analytics

Jui-Sheng Chou [a,*], Chih-Fong Tsai [b], Anh-Duc Pham [a,c], Yu-Hsin Lu [d]

[a] Department of Civil and Construction Engineering, National Taiwan University of Science and Technology, Taiwan
[b] Department of Information Management, National Central University, Taiwan
[c] Faculty of Project Management, The University of Danang, University of Science and Technology, Vietnam
[d] Department of Accounting, Feng Chia University, Taiwan

HIGHLIGHTS

- This comprehensive study used advanced machine learning techniques to predict concrete compressive strength.
- Model performance is evaluated through multi-nation data simulation experiments.
- The prediction accuracy of ensemble technique is superior to that of single learning models.
- This study developed advanced learning approaches for solving civil engineering problems.
- The approach also has potential applications in material sciences.

ARTICLE INFO

Article history:
Received 30 May 2014
Received in revised form 5 September 2014
Accepted 24 September 2014

Keywords:
High performance concrete
Compressive strength
Multi-nation data analysis
Machine learning
Ensemble classifiers
Prediction

ABSTRACT

Machine learning (ML) techniques are increasingly used to simulate the behavior of concrete materials and have become an important research area. The compressive strength of high performance concrete (HPC) is a major civil engineering problem. However, the validity of reported relationships between concrete ingredients and mechanical strength is questionable. This paper provides a comprehensive study using advanced ML techniques to predict the compressive strength of HPC. Specifically, individual and ensemble learning classifiers are constructed from four different base learners, including multilayer perceptron (MLP) neural network, support vector machine (SVM), classification and regression tree (CART), and linear regression (LR). For ensemble models that integrate multiple classifiers, the voting, bagging, and stacking combination methods are considered. The behavior simulation capabilities of these techniques are investigated using concrete data from several countries. The comparison results show that ensemble learning techniques are better than learning techniques used individually to predict HPC compressive strength. Although the two single best learning models are SVM and MLP, the stacking-based ensemble model composed of MLP/CART, SVM, and LR in the first level and SVM in the second level often achieves the best performance measures. This study validates the applicability of ML, voting, bagging, and stacking techniques for simple and efficient simulations of concrete compressive strength.

1. Introduction

An important research problem in materials science is predicting the mechanical properties of construction materials [1]. For many years, the use of high performance concrete (HPC) in various structural applications has markedly increased [2]. Cement materials such as fly ash, blast furnace slag, metakaolin, and silica fume are often used to increase the compressive strength and durability of HPC [3–5]. In terms of concrete mix design and quality control, compressive strength is generally considered the most important quality of HPC. Developing accurate and reliable compressive strength prediction models can save time and costs by providing designers and structural engineers with vital data. Thus, accurate and early prediction of concrete strength is a critical issue in concrete construction.

Concrete compressive strength (CCS) is usually predicted using linear or non-linear regression methods [3,6–8]. The general form of the regression method is

* Corresponding author. Tel.: +886 2 2737 6321; fax: +886 2 2737 6606.
E-mail addresses: jschou@mail.ntust.edu.tw (J.-S. Chou), cftsai@mgt.ncu.edu.tw (C.-F. Tsai), paduc@dut.udn.vn (A.-D. Pham), luyh@fcu.edu.tw (Y.-H. Lu).

http://dx.doi.org/10.1016/j.conbuildmat.2014.09.054
0950-0618/

$$y = f(b_i \cdot x_i) \quad (1)$$

where y, f, b_i and x_i are the CCS, linear or nonlinear function, regression coefficients and concrete attributes, respectively.

However, obtaining an accurate regression equation when using these empirical-based models is difficult. Moreover, several factors that affect the compressive strength of HPC differ from those that affect the compressive strength of conventional concrete. Therefore, regression analysis may be unsuitable for predicting CCS [9].

To compensate for the drawbacks of conventional models, machine learning algorithms (*i.e.*, neural networks, classification and regression tree, linear regression, or support vector machine (SVM)) as baseline models have been applied in evolutionary or hybrid approaches to developing accurate and effective models for predicting CCS [10]. Machine learning, which is a branch of artificial intelligence (AI), can be used not only in knowledge-generation tools, but also in general information-modeling tools for conventional statistical techniques [11]. The ML models approximate the relationships between the inputs and outputs based on a measured set of data.

Recently, the use of ML-based applications has increased in many areas of civil engineering, ranging from engineering design to project planning [11–15]. Other material science problems that have been solved by ML include mixture design, predicting mechanical properties, or fault diagnosis [10,14,16–19]. Particularly, ML-based solutions using learning mechanisms readily available in WEKA software[1] often provide a good alternative approach to solving prediction problems.

For example, Chou et al. proposed several supervised learning models for predicting CCS. Their analytical results indicated that multiple additive regression trees achieve the highest predictive accuracy [20]. Moreover, Yan and Shi reported that SVM was better than other models for predicting elastic modulus of normal and high strength concrete [21]. Notably, artificial neural networks (ANNs) are used to construct mapping functions for predicting CCS [19,22]. Thus, ML techniques such as SVM and ANN are frequently used in prediction models. However, no single model has consistently proven superior.

Conversely, ensemble approaches that combine multiple learning classifiers (or prediction models) have been proposed to improve the performance of single learning techniques [23,24]. The three methods of combining multiple prediction models into a single model are voting, bagging, and stacking combination methods [25–27]. However, a literature review shows no studies that have compared individual models and ensemble learning techniques for predicting the compressive strength of HPC.

Particularly, related works only develop their prediction models based on single statistical or machine learning techniques; it is unknown that whether the prediction models based on ensemble learning techniques can perform better than single ones in the problem of HPC compressive strength prediction.

Therefore, individual and ensemble ML techniques were compared in this study to identify the best model for predicting the mechanical properties of HPC. Specifically, four well-known individual ML techniques are compared: multilayer perceptron neural network, SVM, classification and regression tree, and linear regression. Additionally, the voting, bagging, and stacking integration methods of combining individual models are examined in terms of the mean absolute error (MAE), root mean squared error (RMSE) and mean absolute percentage error (MAPE) *via* a synthesis index. Meanwhile, cross validation method [28] is used to avoid bias in the experimental datasets.

To sum up, the contribution of this paper is two-fold. The first one is to compare the performance of single and ensemble techniques, which is not given thorough investigation in HPC compressive strength prediction. The second one is the identified best prediction model, which provides the lowest error rate, can be used for not only the practical purpose, but also future researches as the baseline prediction model to compare with other advanced models.

The rest of this paper is organized as follows. The study context is introduced by a brief literature review, including studies of CCS prediction and some well-known ML applications. The methodology section then describes individual and ensemble ML schemes and evaluation methods. The modeling experiments section discusses the experimental settings and compares the prediction results among individual and ensemble ML models of HPC compressive strength. Finally, the conclusion summarizes the findings and conclusions.

2. Literature review

The use of computer-aided modeling for predicting the mechanical properties of construction materials is growing [12]. The many prediction techniques proposed so far include empirical models, statistical techniques and artificial intelligence algorithms [3,8,10,13,21,29]. Some linear or non-linear regression analyses have achieved good prediction accuracy. Zain and Abd, for instance, used multivariate power equations to predict the strength of high performance concrete [8]. Similarly, Aciti developed a regression model for estimating concrete strength through non-destructive testing methods and then performed statistical tests to verify the model [7].

However, predicting the behavior of HPC is relatively more difficult than predicting the behavior of conventional concrete. Since the relationship between components and concrete properties is highly non-linear, mathematically modeling the compressive strength of HPC based on available data is difficult [30,31]. Therefore, conventional methods are often unsuitable for predicting concrete compressive strength [9]. Since conventional materials models are inadequate for simulating complex non-linear behaviors and uncertainties, researchers have proposed various AI techniques for enhancing prediction accuracy [9,13,15,20,21]. Boukhatem et al. showed that simulation models, decision support systems, and AI techniques are useful and powerful tools for solving complex problems in concrete technology [12].

Notably, researchers have applied or evaluated the capability of ANNs to predict strength and other concrete behaviors [29,30,32–34]. Ni and Wang, for instance, used multi-layer feed-forward neural networks to predict 28-day CCS based on various factors [32]. Altun et al. further showed that the ANN was superior to regression method in estimating CCS [29]. Yeh also successfully used ANN to predict the slump of concrete with fly ash and blast furnace slag [30]. The use of an adaptive probabilistic neural network for improving accuracy in predicting CCS was studied by Lee et al. [35].

Meanwhile, the SVM has excellent generalization capability when solving non-linear problems. The SVM can also overcome the problem of small sample size. An SVM analysis was used to estimate the temperatures at which concrete structures are damaged by fire [14]. Moreover, Gupta investigated the potential use of SVM for predicting CCS [18] by combining radial basis function with SVM. Yan and Shi used SVM to predict the elastic modulus of normal and high strength concrete. The analytical results showed that the SVM outperformed other models [21].

Similarly, evolutionary algorithm-based methodologies have been used for knowledge discovery. Cheng et al. proposed an advanced hybrid AI model that fused fuzzy logic, weight SVM

[1] http://www.cs.waikato.ac.nz/ml/weka/