Research Article

Application of Boosting Regression Trees to Preliminary Cost Estimation in Building Construction Projects

Yoonseok Shin

Department of Plant and Architectural Engineering, Kyonggi University, Gwanggyosan-ro 154-42, Yeongtong-gu, Suwon, Gyeonggi-do 443-760, Republic of Korea

Correspondence should be addressed to Yoonseok Shin; shinys@kgu.ac.kr

Received 8 October 2014; Revised 11 December 2014; Accepted 7 January 2015

Academic Editor: Rahib H. Abiyev

Copyright © 2015 Yoonseok Shin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Among the recent data mining techniques available, the boosting approach has attracted a great deal of attention because of its effective learning algorithm and strong boundaries in terms of its generalization performance. However, the boosting approach has yet to be used in regression problems within the construction domain, including cost estimations, but has been actively utilized in other domains. Therefore, a boosting regression tree (BRT) is applied to cost estimations at the early stage of a construction project to examine the applicability of the boosting approach to a regression problem within the construction domain. To evaluate the performance of the BRT model, its performance was compared with that of a neural network (NN) model, which has been proven to have a high performance in cost estimation domains. The BRT model has shown results similar to those of the NN model using 234 actual cost datasets of a building construction project. In addition, the BRT model can provide additional information such as the importance plot and structure model, which can support estimators in comprehending the decision making process. Consequently, the boosting approach has potential applicability in preliminary cost estimations in a building construction project.

1. Introduction

In building construction, budgeting, planning, and monitoring for compliance with the client’s available budget, time, and work outstanding are important [1]. The accuracy of the construction cost estimation during the planning stage of a project is a crucial factor in helping the client and contractor with the adequate decision making and for the successful completion of the project [2–5]. However, there is a problem in that it is difficult to quickly and accurately estimate the construction costs at the early stage because the drawings and documentation are generally incomplete [6]. Machine learning approaches can be applied to alleviate this problem. Machine learning has some advantages over the human-crafted rules for data driven works, that is, accurate, automated, fast, customizable, and scalable [7].

Cost estimating approaches using a machine learning technique such as a neural network (NN) or support vector machine (SVM) have received significant attention since the early 1990s for accurately predicting the construction costs under a limited amount of project information. The NN model [1, 8–11] and the SVM model [12–16] were developed for predicting and/or estimating the construction costs. Although applying an NN to construction cost estimations has been very popular and has shown superior accuracy over other competing techniques [2, 4, 17–21], it has several disadvantages, such as a lack of self-learning and a time-consuming rule acquisition process [14]. A SVM, introduced by Vapnik [22], has attracted a great deal of attention because of its capacity for self-learning and high performance in generalization; moreover, it has shown the potential for utilization in construction cost estimations [5, 13, 14, 16, 23, 24]. However, the SVM approach requires a great deal of trial and error to determine a suitable kernel function [14]. Moreover, SVM models have a high level of algorithmic complexity and require extensive amounts of memory [25].

Among the recent machine learning techniques, the boosting approach, which was developed by Freund and Schapire [26], who also introduced the AdaBoost algorithm, has become an important application in machine learning and predicting models [27]. The boosting approach provides an effective learning algorithm and strong boundaries in
terms of the generalization performance [28–31]. Compared with competing techniques used for prediction problems, the performance of the boosting approach is superior to that of both a NN [32] and a SVM [33]. It is also simple, easy to program, and has few parameters to be tuned [31, 34, 35]. Because of these advantages, the boosting approach has been actively utilized in various domains. In the construction domain, some studies have attempted to apply this approach to the classification problem (for predicting a categorical dependent variable), such as the prediction of litigation results [27] and the selection of construction methods [31, 36]. However, there have been no efforts to do so for regression problems (for predicting a continuous dependent variable), such as construction cost estimation.

In this study, the boosting regression tree (BRT) is applied to the cost estimation at the early stage of a construction project to examine the applicability of the boosting approach for a regression problem within the construction domain. The BRT in this study is based on the module of a stochastic gradient boosting tree, which was proposed by Friedman (2002) [37]. It was developed as a novel advance in data mining that extends and improves the regression tree using a stochastic gradient boosting approach. Therefore, it has advantages of not only a boosting approach but also a regression tree, that is, high interpretability, conceptual simplicity, computational efficiency, and so on. The boosting approach can especially adopt other data mining techniques, that is, a NN and SVM, as well as decision tree, as base learner. This feature matches up to the latest trends in the field of fusion of computational intelligence techniques to develop efficient computational models for solving practical problems.

In the next section, the construction cost estimation and its relevant studies are briefly reviewed. In the third section, the theory of a BRT and a cost estimation model using a BRT are both described. In the fourth section, the cost estimation model using a BRT is applied to a dataset from an actual project of a school building construction in Korea and is compared with that of an NN and an SVM. Finally, some concluding remarks and suggestions for further study are presented.

2. Review of Cost Estimation Literature

Raftery [38] categorized the preliminary cost estimation system used in building construction projects into three generations. The first generation of the system was a method from the late 1950s to the late 1960s that utilized the unit price. The second generation of the system, which was developed from the middle of the 1970s, was a statistical method using a regression analysis according to propagating personal computers. The third generation of the system is a knowledge-based artificial intelligence method from the early 1980s. However, based on the third generation, Kim [39] also separated a fourth generation based on machine learning techniques such as a NN and SVM. The author showed an outstanding performance in construction cost estimation, although much remains to be resolved, for example, the complexity of the parameter settings.

We believe that the boosting approach can be a next-generation cost estimation system at the early stage of a construction project. In the prediction problem domain, combining the predictors of several models often results in a model with improved performance. The boosting approach is one such method that has shown great promise. Empirical studies have shown that combining models using the boosting approach produces a more accurate regression model [40]. In addition, the boosting approach can be extensively applied to prediction problems using an aforementioned machine learning technique such as a NN and SVM, as well as decision trees [27]. However, the boosting approach has never been used in regression problems of the construction domain, including cost estimations, but has been actively utilized in other domains, such as remote aboveground biomass retrieval [41], air pollution prediction [42], software effort estimation [43], soil bulk density prediction [44], and Sirex noctilio prediction [45]. In this study, we examine the applicability of a BRT for estimating the costs in the construction domain.

3. Boosting Regression Trees

Because of the abundance of exploratory tools, each having its own pros and cons, a difficult problem arises in selecting the best tool. Therefore, it would be beneficial to try to combine their strengths to create an even more powerful tool. To a certain extent, this idea has been implemented in a new family of regression algorithms referred to under the general term of “boosting.” Boosting is an ensemble learning method for improving the predictive performance of a regression procedure, such as the use of a decision tree [46]. As shown in Figure 1, the method attempts to boost the accuracy of any given learning algorithm by fitting a series of models, each having a low error rate, and then combining them into
Stochastic gradient boosting is a novel advance to the boosting approach proposed by Friedman [37] at Stanford University. Of the previous studies [26, 49–51] related to boosting for regression problems, only Breiman [50] alludes to involving the optimization of a regression loss function as part of the boosting algorithm. Friedman [52] proposed using the connection between boosting and optimization, that is, the gradient boost algorithm. Friedman [37] then showed that a simple subsampling trick can greatly improve the predictive performance of stochastic gradient boost algorithms while simultaneously reducing their computational time.

The stochastic gradient boost algorithm proposed by Friedman [37] uses regression trees as the basis functions. Thus, this boosting regression tree (BRT) involves generating a sequence of trees, each grown on the residuals of the previous tree [46]. Prediction is accomplished by weighting a sequence of trees, each grown on the residuals of the previous tree [46]. Prediction is accomplished by weighting the ensemble outputs of all regression trees, as shown in Figure 2 [53]. Therefore, this BRT model inherits almost all of the advantages of tree-based models, while overcoming their primary disadvantages, that is, inaccuracies [54].

In these algorithms, the BRT approximates the function $F(x)$ as an additive expansion of the base learner (i.e., a small tree) [43]:

$$F(x) = F_0(x) + \beta_1 F_1(x) + \beta_2 F_2(x) + \cdots + \beta_m F_m(x). \quad (1)$$

A single base learner does not make sufficient prediction using the training data, even when the best training data are used. It can boost the prediction performance using a series of base learners with the lowest residuals.

Technically, BRT employs an iterative algorithm, where, at each iteration $m$, a new regression tree $h(x; R_{lm})$ partitions the $x$-space into $L$-disjoint regions $\{R_{lm}\}$ and predicts a separate constant value in each region $[54]$:

$$h(x; R_{lm}) = \sum_{l=1}^{L} \bar{y}_{lm} | (x \in R_{lm}). \quad (2)$$

Here $\bar{y}_{lm} = \text{mean}_{x \in R_{lm}}(\bar{y}_{lm})$ is the mean of pseudo-residuals (3) in each region $R_{lm}$ induced at the $m$th iteration [37, 54]:

$$\bar{y}_{lm} = \left. \frac{\partial \Psi(y_i, F(x_i))}{\partial F(x_i)} \right|_{F(x) = F_{m-1}(x)}. \quad (3)$$

The current approximation $F_{m-1}(x)$ is then separately updated in each corresponding region [37, 54]:

$$F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_m | (x \in R_{lm}), \quad (4)$$

where

$$\gamma_m = \arg \min_{\gamma} \sum_{x_i \in R_{lm}} \Psi(y_i, F_{m-1}(x_i) + \gamma). \quad (5)$$

The “shrinkage” parameter $\nu$ controls the learning rate of the procedure.

This leads to the following BRT algorithm for generalized boosting of regression trees [37].

1. Initialize $F(x), F_0(x) = \arg \min_{\gamma} \sum_{i=1}^{N} \Psi(y_i, \gamma)$.
2. For $m = 1$ to $M$ do
3. Select a subset randomly from the full training dataset,

$$\{\pi(i)\}^N_i = \text{rand_perm} \{i\}^N.$$

4. Fit the base learner,

$$\bar{y}_{\pi(i)m} = \left[ \frac{\partial \Psi(y_{\pi(i)m}, F(x_{\pi(i)})})}{\partial F(x)} \right]_{F(x) = F_{m-1}(x)}, \quad i = 1, N. \quad (7)$$

5. Compute the model update for the current iteration,

$$\{R_{lm}\} = \text{L−terminal node tree}\left(\{\bar{y}_{\pi(i)m}, x_{\pi(i)}\}\right)^N_i. \quad (8)$$

6. Choose a gradient descent step size as,

$$\gamma_m = \arg \min_{\gamma} \sum_{x_i \in R_{lm}} \Psi(y_i, F_{m-1}(x_{\pi(i)}) + \gamma). \quad (9)$$

7. Update the estimate of $F(x)$ as,

$$F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_m | (x \in R_{lm}). \quad (10)$$

8. end For.

There are specific algorithms for several loss criteria including least squares: $\psi(y, F) = (y - F)^2$, least-absolute deviation: $\psi(y, F) = |y - F|$, and Huber-M: $\psi(y, F) = (y - F)^2 \cdot |(y - F) < \delta| + 2\delta(y - F - \delta/2) \cdot |(y - F) > \delta| [37]$. The BRT applied in this study adopts the least squares for loss criteria as shown in Figure 3.

4. Application

4.1. Determining Factors Affecting Construction Cost Estimation. In general, the estimation accuracy in a building project is correlated with the amount of project information available regarding the building size, location, number of stories, and so forth [55]. In this study, the factors used for estimating the construction costs are determined in two steps. First, a
### Table 1: Factors in construction cost estimation.

<table>
<thead>
<tr>
<th>Description</th>
<th>Min.</th>
<th>Max.</th>
<th>Average</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Budget</td>
<td>(1) BTL</td>
<td>Nominal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) National finance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School levels</td>
<td>(1) Elementary</td>
<td>Nominal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Middle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Land acquisition</td>
<td>(1) Existing</td>
<td>Nominal</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2) Building lots</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3) Green belts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class number</td>
<td>12</td>
<td>48</td>
<td>31</td>
<td>Numerical</td>
</tr>
<tr>
<td>Building area (m²)</td>
<td>1,204</td>
<td>3,863</td>
<td>2,694</td>
<td>Numerical</td>
</tr>
<tr>
<td>Gross floor area (m²)</td>
<td>4,925</td>
<td>12,710</td>
<td>9,656</td>
<td>Numerical</td>
</tr>
<tr>
<td>Storey</td>
<td>3</td>
<td>7</td>
<td>4.7</td>
<td>Numerical</td>
</tr>
<tr>
<td>Basement floor (storey)</td>
<td>0</td>
<td>2</td>
<td>0.5</td>
<td>Numerical</td>
</tr>
<tr>
<td>Floor Height (m)</td>
<td>3.3</td>
<td>3.6</td>
<td>3.5</td>
<td>Numerical</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total construction cost (thousand KRW)</td>
<td>4,334,369</td>
<td>14,344,867</td>
<td>8,288,008</td>
<td>Numerical</td>
</tr>
</tbody>
</table>

Factors (i.e., input variables) were selected for this study, as shown in Table 1.

#### 4.2 Data Collection.

Data were collected from 234 completed school building projects executed by general contractors from 2004 to 2007 in Gyeonggi Province, Korea. These cost data were only the direct costs of different school buildings, such as elementary, middle, and high schools, without a markup as shown in Figure 4. According to the construction year, the total construction costs were converted using the Korean building cost index (BCI); that is, the collected cost data were multiplied by the BCI of the base year of 2005 (BCI = 1.00). The collected cost data of 217 school buildings were randomly divided into 30 test datasets and 204 training datasets.

#### 4.3 Applying BRT to Construction Cost Estimation.

In this study, the construction cost estimation model using a BRT was tested through application to real building construction projects. The construction costs were estimated using the BRT as follows. (1) The regression function $\hat{F}(x)$ was trained using training data. In the dataset, the budget, school levels, gross floor area, and so on were allocated to each $x_i$ of the training set. Each result, that is, the actual cost, was allocated to $y_i$. (2) After the training was completed according to the parameters such as the learning (shrinkage) rate, the number of additive trees, and the maximum and minimum number of levels, the series of trees $\hat{F}(x)$ which maps $x$ to $y$ of training data set ($y_i, x_i$) with minimized loss function $\Psi(y_i, \hat{F}(x_i))$ was found. (3) The expected value of $\hat{F}(x)$, that is, the expected cost, was calculated for a new test dataset ($y_j, x_j$). The construction cost estimation model proposed in this study was constructed using “STATISTICA Release 7.” STATISTICA employs an implementation method usually

![Figure 3: Training results of BRT.](image-url)
referred to as a stochastic gradient boosting tree by Friedman (2002, 2001) [37, 52], also known as TreeNet (Salford Systems, Inc.) or MART (Jerill, Inc.). In this software, a stochastic gradient boosting tree is used for regression problems to predict a continuous dependent variable [57]. To operate a boosting procedure in STATISTICA, the parameter settings, that is, the learning rate, the number of additive trees, the proportion of subsampling, and so forth, are required. Firstly, the learning rate was set as 0.1. It was found that small values, that is, values under 0.1, lead to much better results in terms of the prediction error [52]. We empirically obtained the other parameters, which are shown in Figure 5. As a result, the training result of the BRT showed that the optimal number of additive trees is 183 and the maximum size of tree is 5, as shown in Figure 3.

4.4. Performance Evaluation. In general, the cost estimation performance can be measured based on the relationship between the estimated and actual costs [56]. In this study, the performance was measured using the Mean Absolute Error Rates (MAERs), which were calculated using

$$\text{MAERs} = \frac{1}{n} \sum \left| \left( \frac{C_e - C_a}{C_a} \right) \times 100 \right|$$

where $C_e$ is the estimated construction costs by model application, $C_a$ is the actual construction costs collected, and $n$ is the number of test datasets.

To verify the performance of the BRT model, the same cases were applied to a model based on a NN and the results compared. We chose the NN model because it showed a superior performance in terms of cost estimation accuracy in previous studies [2, 5, 14]. “STATISTICA Release 7” was also used to construct the NN model in this study. To construct a model using a NN, the optimal parameters have to be selected beforehand, that is, the number of hidden neurons, the momentum, and the learning rate for the NN. Herein, we determined the values from repeated experiments.

5. Results and Discussion

5.1. Results of Evaluation. The results from the 30 test datasets using a BRT and a NN are summarized in Tables 2 and 3. The results from the BRT model had MAERs of 5.80 with 20% of the estimates within 2.5% of the actual error rate, while 80% were within 10%. The NN model had MAERs of 6.05 with 10% of the estimates within 2.5% of the actual error rate, while 93.3% were within 10%. In addition, the standard deviations of the NN and BRT models are 3.192 and 3.980, respectively, as shown in Table 4.

The MAERs of the two results were then compared using a $t$-test analysis. The MAERs of the two results are statistically similar, although there are differences between them. As the null hypothesis, the MAERs of the two results are all equal ($H_0: \mu_1 = \mu_2$). The $t$-value is 0.263 and the $P$ value is 0.793 (>0.05). Thus, the null hypothesis is accepted. This analysis shows that the MAERs of the two results are statistically similar.

The BRT model provided comprehensible information regarding the new cases to be predicted, which is an advantage inherent to a decision tree. Initially, the importance of each dependent variable to cost estimation was provided, as shown in Figure 6. These values indicate the importance...
of each variable for the construction cost estimation in the model. Finally, the tree structures in the model were provided as shown in Figure 7. This shows the estimation rules, such as the applied variables and their influence on the proposed model. Thus, an intuitive understanding of the whole structure of the model is possible.

5.2. Discussion of Results. This study was conducted using 234 school building construction projects. In addition, 30 of these projects were used for testing. In terms of the estimation accuracy, the BRT model showed slightly better results than the NN model, with MAERs of 5.80 and 6.05, respectively. In terms of the construction cost estimation, it is difficult to conclude that the performance of the BRT model is superior to that of the NN model because the gap between the two is not statistically different. However, even the similar performance of the BRT model is notable because the NN model has proven its superior performance in terms of cost estimation accuracy in previous studies. Similarly, in predicting the software project effort, Elish [43] compared the estimation accuracy of neural network, linear regression, support vector regression (SVR), and BRT. Consequently, BRT outperformed the other techniques in terms of the estimation performance that has been also achieved by SVR. These results mean that the BRT has remarkable performance in regression problems as well as classification one. Moreover, the BRT model provided additional information, that is, an importance plot and structure model, which helps the estimator comprehend the decision making process intuitively.

Consequently, these results reveal that a BRT, which is a new AI approach in the field of construction, has potential applicability in preliminary cost estimations. It can assist estimators in avoiding serious errors in predicting the construction costs when only limited information is available during the early stages of a building construction project. Moreover, a BRT has a large utilization possibility because the boosting approach can employ existing AI techniques such as a NN and SVM, along with decision trees, as base learners during the boosting procedure.
Table 3: Cost estimation results of each test set.

<table>
<thead>
<tr>
<th>Number</th>
<th>Historical cost (1,000 KRW)</th>
<th>Neural networks Predicted cost (1,000 KRW)</th>
<th>Error rate (%)</th>
<th>Boosting regression tree Predicted cost (1,000 KRW)</th>
<th>Error rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6,809,450</td>
<td>7,704,034</td>
<td>13.14</td>
<td>7,206,795</td>
<td>5.84</td>
</tr>
<tr>
<td>2</td>
<td>9,351,716</td>
<td>10,015,906</td>
<td>7.10</td>
<td>9,805,656</td>
<td>4.85</td>
</tr>
<tr>
<td>3</td>
<td>6,656,230</td>
<td>7,251,317</td>
<td>8.94</td>
<td>6,322,112</td>
<td>5.02</td>
</tr>
<tr>
<td>4</td>
<td>7,119,470</td>
<td>7,128,513</td>
<td>0.13</td>
<td>7,418,373</td>
<td>4.20</td>
</tr>
<tr>
<td>5</td>
<td>7,304,747</td>
<td>7,978,990</td>
<td>9.23</td>
<td>7,349,178</td>
<td>0.61</td>
</tr>
<tr>
<td>6</td>
<td>9,729,392</td>
<td>9,516,946</td>
<td>2.18</td>
<td>9,259,162</td>
<td>4.83</td>
</tr>
<tr>
<td>7</td>
<td>10,801,826</td>
<td>9,817,999</td>
<td>9.11</td>
<td>9,682,119</td>
<td>10.37</td>
</tr>
<tr>
<td>8</td>
<td>7,944,318</td>
<td>7,246,763</td>
<td>8.78</td>
<td>7,136,773</td>
<td>10.17</td>
</tr>
<tr>
<td>9</td>
<td>10,879,004</td>
<td>10,136,431</td>
<td>6.83</td>
<td>10,572,777</td>
<td>2.81</td>
</tr>
<tr>
<td>10</td>
<td>7,552,814</td>
<td>7,764,300</td>
<td>2.80</td>
<td>7,683,295</td>
<td>1.73</td>
</tr>
<tr>
<td>11</td>
<td>8,845,099</td>
<td>8,558,536</td>
<td>3.24</td>
<td>8,370,497</td>
<td>5.37</td>
</tr>
<tr>
<td>12</td>
<td>10,690,800</td>
<td>10,001,503</td>
<td>6.45</td>
<td>10,015,284</td>
<td>6.32</td>
</tr>
<tr>
<td>13</td>
<td>8,694,721</td>
<td>8,258,452</td>
<td>5.02</td>
<td>8,446,796</td>
<td>2.85</td>
</tr>
<tr>
<td>14</td>
<td>6,582,636</td>
<td>6,810,406</td>
<td>3.46</td>
<td>6,954,507</td>
<td>5.65</td>
</tr>
<tr>
<td>15</td>
<td>7,583,680</td>
<td>8,312,216</td>
<td>9.61</td>
<td>8,194,292</td>
<td>8.05</td>
</tr>
<tr>
<td>16</td>
<td>7,099,220</td>
<td>7,955,966</td>
<td>12.07</td>
<td>8,292,381</td>
<td>16.81</td>
</tr>
<tr>
<td>17</td>
<td>8,145,147</td>
<td>8,604,444</td>
<td>5.64</td>
<td>8,522,009</td>
<td>4.63</td>
</tr>
<tr>
<td>18</td>
<td>8,652,810</td>
<td>7,853,765</td>
<td>9.23</td>
<td>8,270,169</td>
<td>4.42</td>
</tr>
<tr>
<td>19</td>
<td>10,527,278</td>
<td>10,040,039</td>
<td>4.63</td>
<td>9,611,194</td>
<td>8.70</td>
</tr>
<tr>
<td>20</td>
<td>6,679,924</td>
<td>6,467,344</td>
<td>3.18</td>
<td>7,397,923</td>
<td>10.75</td>
</tr>
<tr>
<td>21</td>
<td>8,383,830</td>
<td>9,203,887</td>
<td>9.78</td>
<td>8,487,286</td>
<td>1.23</td>
</tr>
<tr>
<td>22</td>
<td>7,298,932</td>
<td>8,018,225</td>
<td>9.85</td>
<td>8,294,895</td>
<td>13.65</td>
</tr>
<tr>
<td>23</td>
<td>7,505,428</td>
<td>7,749,053</td>
<td>3.25</td>
<td>7,967,265</td>
<td>6.15</td>
</tr>
<tr>
<td>24</td>
<td>7,710,921</td>
<td>7,622,053</td>
<td>1.15</td>
<td>7,795,563</td>
<td>1.10</td>
</tr>
<tr>
<td>25</td>
<td>6,196,652</td>
<td>6,503,022</td>
<td>4.94</td>
<td>5,940,634</td>
<td>4.13</td>
</tr>
<tr>
<td>26</td>
<td>8,897,861</td>
<td>8,554,455</td>
<td>3.86</td>
<td>8,714,123</td>
<td>2.06</td>
</tr>
<tr>
<td>27</td>
<td>7,840,787</td>
<td>8,535,617</td>
<td>8.86</td>
<td>8,863,975</td>
<td>13.05</td>
</tr>
<tr>
<td>28</td>
<td>8,023,067</td>
<td>7,666,898</td>
<td>4.44</td>
<td>6,900,068</td>
<td>14.00</td>
</tr>
<tr>
<td>29</td>
<td>7,495,213</td>
<td>7,270,806</td>
<td>2.99</td>
<td>7,695,613</td>
<td>2.67</td>
</tr>
<tr>
<td>30</td>
<td>7,653,005</td>
<td>8,003,292</td>
<td>4.58</td>
<td>7,775,139</td>
<td>1.60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MAERS</th>
<th>Std, deviation</th>
<th>Std, error</th>
<th>95% confidence interval of the MAERS Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>6.045</td>
<td>3.192</td>
<td>0.583</td>
<td>2.542</td>
</tr>
<tr>
<td>BRT</td>
<td>5.800</td>
<td>3.980</td>
<td>0.727</td>
<td>3.170</td>
</tr>
</tbody>
</table>

6. Conclusion

This study applied a BRT to construction cost estimation, that is, the regression problem, to examine the applicability of the boosting approach to a regression problem in the construction domain. To evaluate the performance of the BRT model, its performance was compared with that of an NN model, which had previously proven its high performance capability in the cost estimation domains. The BRT model showed similar results when using 234 actual cost datasets of a building construction project in Korea. Moreover, the BRT model can provide additional information regarding the variables to support estimators in comprehending the decision making process. These results demonstrated that the BRT has dual advantages of boosting and decision trees. The boosting approach has great potential to be a leading technique in next generation construction cost estimation systems.

In this study, an examination using a relatively small dataset and number of variables was carried out on the performance of a BRT for construction cost estimation. Although
both models performed satisfactorily, further detailed experiments and analyses regarding the quality of the collected data are necessary to utilize the proposed model for an actual project.

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work was supported by Kyonggi University Research Grant 2012.

References


39. G. H. Kim, Construction cost prediction system based on artificial intelligence at the project planning stage [Ph.D. thesis], Korea University, Seoul, Republic of Korea, 2004.


Submit your manuscripts at http://www.hindawi.com