Abstract

The fate of many construction projects is determined using preliminary project cost estimates. These estimates play a key role during the conceptual phase of projects; as in many cases they are the primary element used to decide their viability. The lack of information and the high levels of uncertainty with what will be done in the project, during the conceptual phase, make it infeasible to have reliable building information models that could be used to generate quantity takeoffs for preliminary cost estimates (known in the industry as 5D BIM), in line with a Level 2 BIM maturity. This paper presents a way to combine artificial intelligence (case-based reasoning and neural networks), with traditional techniques (regression analysis), to develop accurate estimates of the resources needed in a project (e.g., construction material quantities). These estimates of resources can then be coupled with unit cost information to make preliminary resource-based cost estimates. The clear division between the technical and financial components of such an estimate give improved decision support to project managers and decision makers. This enhances the tracking and control mechanisms which could be used to check the estimates prepared in subsequent project phases. The combination of case-based reasoning with regression analysis and the use of neural networks has shown an improved performance in the estimation of the amount construction material quantities. The proposed combination was used to estimate the amount of concrete, reinforcement, and structural steel required for the construction of tall-frame structures. The results show lower errors (overall mean absolute percentage error-MAPE) for the combined models (2.55%) when compared to the regression models (12.01%), neural network models (5.84%), and case-based reasoning models (9.30%). This type of estimates will help keep construction projects on schedule and on budget.

Keywords: case-based reasoning; hybrid estimation models; neural networks; regression analysis; preliminary estimates of resources.

1. Introduction

In the majority of projects requiring the use of significant amounts of capital, both financial and labor, the norm is that projects have a tendency to costs more than originally planned. Studies ([1], [2]) have shown that using current cost estimating practices over 80% of the projects investigated from different industries (e.g., oil and gas, chemical, power and utilities, mining and metals, EPC contractors, manufacturing, telecommunications, government and defense, transportation) were over budget with actual cost exceeding 40% of the original estimated amounts. This is not limited to a particular sector or country, but a general phenomenon that occurs in most industries and in most nations. In some cases if the true value of the construction costs were known at the approval stage, the project may not have been approved. In addition, cost estimates have not improved and cost overruns not decreased over the past 70 years ([3]).

Preliminary cost estimates are the first thoughtful efforts to predict the cost of a project and they are crucial to the initial decision-making process for the construction of projects. These estimates heavily influence the fate of many projects, yet they have many limitations. Estimators at these early stages of a project’s life cycle are typically provided with very little information about the project, limited scope definition, and very little time to prepare this kind of estimates.
Preliminary cost estimates are typically carried out to support project evaluations, engineering designs, cost budgeting and cost management. Many methods have been developed in various industries to evaluate costs at pre-design stage ([4], [5]). In most cases, construction and infrastructure managers make estimates based heavily on their expert knowledge, missing the full use of the information that they possess and using a great amount of time and effort. In addition, cost estimating is a tedious and time-consuming quantification process, prone to human error which tends to propagate inaccuracies into the different line items of an estimate.

2. State of the art

The estimation models used can be very rudimentary or highly complex and proprietary. Many techniques for the estimation of construction cost have been employed by researchers and practitioners. While most practitioners use models based on regression analysis (RA) or make estimates based on their experience, some researchers and sophisticated owners and contractors use artificial intelligence (AI) techniques, such as neural networks (NNs), and case-based reasoning (CBR). In some cases, combinations of techniques have been used.

The concept of combining estimates to improve the estimation in not new. [6] suggested combining estimates through regression. However, it was the work of [7] and [8] that provided the initial impetus to the development of a theory in the combination of estimates. Combining estimates is especially useful when it is not clear which method would provide more accurate ones ([9]).

The purpose of combining estimates is to use each model’s unique features to capture different patterns or features in the data set ([10]) and by combining estimates from different models, the accuracy can often be improved over the individual estimate ([11]). In the construction industry, the concept of combining different techniques has been explored by different researchers to develop cost estimates (Table 8).

Table 8: Combination of different techniques to develop cost estimates in the construction industry

<table>
<thead>
<tr>
<th>Source</th>
<th>Estimate for</th>
<th>Combination of</th>
</tr>
</thead>
<tbody>
<tr>
<td>[12] Choi et al. (2013)</td>
<td>Cost of public roads in Korea</td>
<td>RST, CBR, and GA</td>
</tr>
</tbody>
</table>

Legend: NN: Neural network; GA: Genetic algorithm; CBR: Case-based reasoning; NFS: Neuro Fuzzy System; AHP: Analytic Hierarchy Process; MRA: Multiple Regression Analysis; RST: Rough Set Theory

Although the combination of techniques to develop cost estimates provided good results, these estimates typically provide a single monetary value and they lack the information needed to make meaningful comparisons with other, more accurate, estimates developed during the project’s life cycle. This also affects the information used in the different control tools available to project managers and limits applications (lessons learned) for future projects.

The construction industry should take full advantage of new techniques and borrow expertise from other fields to implement methods such as AI, and adopt newer technologies, such as digital modeling software that allows computer generation of n-dimensional models (generally associated with Building Information Modeling) to improve the accuracy and reliability of preliminary cost estimates. To this end, a way to assist the estimator in the preparation of reliable and traceable estimates made during the project conception phase of construction projects is presented in this paper that shows a way to develop preliminary estimates for resources (e.g., construction material quantities-CMQs) combining AI approaches (e.g., NNs and CBR) and traditional techniques (e.g., RA).

As done by other researchers (e.g., [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31]) the preliminary estimates were made for the amount of resources (e.g., CMQs) to be used in a given project so that there is a clear separation between technical estimates (e.g., quantities) and market fluctuations (e.g., cost of materials and labor).

3. Combination of AI approaches and traditional techniques

In the study presented in this paper, RA, NNs, and CBR were used to make preliminary CMQ estimates in a systematic, reliable, and accurate manner. Multiple regression was used to develop regression models using a
particular type of nonlinear relationship (Equation (1)), known as the constant elasticity or multiplicative relationship ([32]).

\[
Y = \left( X_1^{\beta_0} X_2^{\beta_1} \ldots X_n^{\beta_n} \right) e^{\left( \beta_0 + \sum_{j=1}^{m} \beta_{n+j} X_{n+j} \right)}
\]  

(1)

Where,
\[
y: \text{Output from the regression equation} \\
\beta_0: \text{Constant} \\
\beta_1 \rightarrow \beta_n: \text{Coefficients for continuous variables} \\
\beta_{n+1} \rightarrow \beta_{n+m}: \text{Coefficients for categorical variables} \\
X_1 \rightarrow X_n: \text{Continuous independent variables} \\
X_{(n+1)} \rightarrow X_{(n+m)}: \text{Categorical independent variables} \\
e: \text{Euler's number} \\
SEE: \text{Standard error of the estimate}
\]

The NN models developed were fully connected feed-forward network with three layers (one hidden layer) using the back-propagation supervised learning algorithm. The independent variables from the regression models were used as the neurons in the input layer.

CBR was used with its four phases, namely retrieve, reuse, revise and retain [33]. The retrieve and revise (i.e., adaptation) phases were modified using information from the regression models.

The combination of the AI approaches and traditional techniques is summarized in Figure 10. The coefficients from the developed regression models for a given CMQ were used by the retrieval and adaptation phases in CBR, and the independent variables (IVs) were used to determine the neurons in the input layer of the NN models, which provided better results for complex and highly non-linear cases. These models were also used to perform direct estimations in the cases where similar structures were not found using CBR (i.e., CBR cannot be used).

![Figure 10: Interaction among the different techniques used to make CMQ estimates](image)

Three basic concepts were kept in mind when developing a way to develop CMQ estimates, they were: learning, adjusting and estimating. These concepts were integrated by combining RA and AI.
3.1. Learning from historical data

The combined techniques made use of the CMQs of existing structures to estimate the CMQs of target structures, i.e., the ones for which the CMQs are to be estimated. The target structures were compared to the existing ones by determining their similarity based on the values of numerous parameters. The most similar structures were used as the basis of the estimate for the target structures.

This was done by using CBR, in particular during the retrieval process. The proposed retrieval process used the city-block distance with the adjusted unstandardized coefficients obtained from RA as shown in [34].

3.2. Adaptation of proposed results

The CMQs of the similar existing structures were adapted. This adaptation accounted for the differences between the target and the existing structures by adjusting the proposed estimation of CMQs for the target structure when applying the selected estimation model. This adaptation was done by incorporating a systematic process in which the regression models were used as the basis of the adaptation process. More information about this adaptation process can be found in [35].

3.3. Direct estimation

Estimates of the CMQs, if no similar structures were found, was done by directly using either the regression of the NN model developed. More information about the development and evaluation of estimation models using RA and NN can be found in [31].

4. Estimate CMQs for a new project

The CMQs for the structures in the target project were estimated using a combination of RA, NN, and CBR, which interacted with each other to provide accurate results. The coefficients from the developed regression models for a given CMQ were used by the retrieval and adaptation phases in CBR, and the independent variables (IVs) were used to determine the neurons in the input layer of the NN models which provided better results for complex and highly non-linear cases. These models were used to perform direct estimations in the cases where there were not similar structures identified using CBR (Figure 10).

Using CBR, similar structures were identified and used as the bases for the new estimate. This provided the transfer of knowledge (i.e., learning from historical data) to this process. The CMQs from the similar existing structures were adjusted to take into consideration the differences between the values of the input parameters from the target structure and the similar existing structures. In the cases where CBR was not applicable, either RA models or NN models were employed to complete the estimate of CMQs for the new project. Throughout this phase the “bases for the estimated CMQs” (e.g., the parameters used and the assumptions relied upon in the estimating process) was tracked and documented.

4.1. Implementation

Models were developed using RA, NN, and CBR to estimate the amount of CMQs required in the construction of the upper structure and foundation of tall-frame structures. The information used to develop the models was limited to the information that would be readily available to the estimator during the early stages of a project. The tall-frame structures had a rectangular area and were built of reinforced concrete and structural steel. A total of 148 tall-frame structures with the characteristics shown in Table 9 were used for model development and model testing. The set was randomly divided into a set of 118 for testing and 30 for testing the model.

### Table 9: Characteristics of tall-frame structures used to estimate CMQs

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Max</th>
<th>Min</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height (m)</td>
<td>158</td>
<td>94</td>
<td>134</td>
</tr>
<tr>
<td>Footprint (m²)</td>
<td>150</td>
<td>35</td>
<td>84</td>
</tr>
<tr>
<td>Wind speed (m/s)</td>
<td>54</td>
<td>31</td>
<td>41</td>
</tr>
<tr>
<td>Ground acc. (S₁ₓg)</td>
<td>1.32</td>
<td>1.02</td>
<td>1.16</td>
</tr>
<tr>
<td>Soil BC (t/m²)</td>
<td>65</td>
<td>20</td>
<td>37</td>
</tr>
<tr>
<td>Concrete (m³)</td>
<td>11,560</td>
<td>2,814</td>
<td>5,288</td>
</tr>
<tr>
<td>Rebar (t)</td>
<td>1,879</td>
<td>442</td>
<td>842</td>
</tr>
<tr>
<td>Structural steel (t)</td>
<td>734</td>
<td>367</td>
<td>542</td>
</tr>
</tbody>
</table>
The regression and NN models for each CMQ are summarized in Table 10 and Table 11 respectively. Table 10 shows the coefficients from the regression models for the different materials estimated. Table 11 shows the weights between the input and hidden layer and hidden layer and output layer for the different NN models developed for each CMQ.

Table 10: Regression models for the estimation of CMQs of tall-frame structures

<table>
<thead>
<tr>
<th>Concrete (m³)</th>
<th>Reinforcement (tons)</th>
<th>Structural steel (tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variables</strong></td>
<td><strong>Unstandardized Coefficients</strong></td>
<td><strong>Standardized Coefficients</strong></td>
</tr>
<tr>
<td>(Constant)</td>
<td>β</td>
<td>Std. Error</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.27E+00</td>
<td>8.00E-02</td>
</tr>
<tr>
<td>Footprint (m²)</td>
<td>3.23E-01</td>
<td>5.00E-03</td>
</tr>
<tr>
<td>Wind speed (m/s)</td>
<td>-6.00E-03</td>
<td>6.00E-03</td>
</tr>
<tr>
<td>Ground acc. (S, xg)</td>
<td>-1.83E-01</td>
<td>2.00E-03</td>
</tr>
<tr>
<td>Soil BC (t/m³)</td>
<td>1.50E-01</td>
<td>5.00E-03</td>
</tr>
</tbody>
</table>

Table 11: NN models for the estimation of CMQs of tall-frame structures

<table>
<thead>
<tr>
<th>Concrete (m³)</th>
<th>Reinforcement (tons)</th>
<th>Structural steel (tons)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Input</strong></td>
<td><strong>Output Layer</strong></td>
<td>**Hidden Layer (Wₓ)</td>
</tr>
<tr>
<td>H₁: Height (m)</td>
<td>8.20E-02</td>
<td>-6.00E-01</td>
</tr>
<tr>
<td>H₂: Footprint (m²)</td>
<td>1.72E+01</td>
<td>2.30E-01</td>
</tr>
<tr>
<td>H₃: Wind speed (m/s)</td>
<td>-3.00E-03</td>
<td>3.00E-03</td>
</tr>
<tr>
<td>H₄: Ground acc. (S, xg)</td>
<td>-1.10E-02</td>
<td>-1.72E-01</td>
</tr>
<tr>
<td>Bias 1</td>
<td>-1.95E-01</td>
<td>5.90E-02</td>
</tr>
</tbody>
</table>

4.2. Results

The results obtained for the estimation of the amount of concrete, reinforcement, and structural steel using the combination of CBR, RA and NN models showed a close agreement between the actual and estimated amounts. An example is shown in Figure 11 for the amount of structural steel as can be seen by the closeness between the actual and estimated amounts.

Figure 11: Actual and estimated amount of structural steel for the construction of tall-frame structures
The mean absolute percentage error (MAPE) of the estimates using the combination of CBR, RA, and NN was lower than the MAPE for the stand-alone models (i.e., without combination) for all CMQs estimated (Table 12).

<table>
<thead>
<tr>
<th>CMQ</th>
<th>Combined (CBR, RA, NN)</th>
<th>RA</th>
<th>NN</th>
<th>CBR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concrete (m³)</td>
<td>2.73%</td>
<td>12.58%</td>
<td>8.65%</td>
<td>9.09%</td>
</tr>
<tr>
<td>Reinforcement (tons)</td>
<td>2.60%</td>
<td>11.39%</td>
<td>4.86%</td>
<td>9.56%</td>
</tr>
<tr>
<td>Structural steel (tons)</td>
<td>2.31%</td>
<td>12.07%</td>
<td>4.01%</td>
<td>9.24%</td>
</tr>
<tr>
<td>Overall MAPE</td>
<td>2.55%</td>
<td>12.01%</td>
<td>5.84%</td>
<td>9.30%</td>
</tr>
</tbody>
</table>

5. Conclusion

The combination of AI approaches and traditional techniques has proven to be an appropriate way to develop estimates of the resources needed in a construction, such as the amount of CMQs required. These estimates of resources can then be coupled with cost information to get preliminary resource-based cost estimates with a clear division between the technical and financial components. This type of estimates give better decision support to project managers and decision makers, and subsequently, enhance the tracking and control mechanisms used in subsequent project phases. The combination of RA, with NN and CBR has shown an improved performance in the estimation of CMQs of tall-frame structures when compared to the performance of these techniques alone as shown by the overall MAPE for the combined models (2.55%), regression models (12.01%), NN models (5.84%), and CBR models (9.30%). Although the implementation has been made for tall-frame structures, this approach could be suited for other structure types and other resources.

The three concepts of learning, adjusting and estimating have been successfully integrated by combining RA and AI by using the coefficients from the developed regression models during the retrieval and adaptation phases in CBR, and the independent variables from the regression models as the neurons in the input layer of the NN models. Even in cases where CBR was not able to find similar existing cases, estimations were also possible by using the appropriate regression or NN model.

This idea of combination of techniques to develop resource-based cost estimates has a wide ranging potential to be used for other types of infrastructure and many other types of projects. The development of preliminary estimates in this way will help estimators to make better estimates, and armed with this information, allow managers to make better decisions of what should be done and when.

Acknowledgements

The authors would like to thank the Project Management & Engineering function at LafargeHolcim, especially to Mr. Rudy Blum and Mr. Roberto Nores for their collaboration and support during this research. Special acknowledgement is given to Prof. Dr. Dilum Fernando from The University of Queensland for his guidance during the development of this work.

References


