Expert Systems with Applications 39 (2012) 5214–5222

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

MRA-based revised CBR model for cost prediction in the early stage of construction projects

RunZhi Jin [a], KyuMan Cho [b,*], ChangTaek Hyun [a], MyungJin Son [a]

[a] Department of Architectural Engineering, University of Seoul, Seoul, South Korea
[b] Division of Construction Engineering and Management, Purdue University, West Lafayette, IN, United States

ARTICLE INFO

Keywords:
Case-Based Reasoning
Multiple Regression Analysis
Cost prediction

ABSTRACT

Accurate prediction of construction cost in the initial phase of a construction project is critical to the success of the project. Accordingly, many researchers have proposed various methodologies for predicting the cost in the initial phase with the use of limited information. This study was aimed at improving the prediction performance of a cost prediction model based on the Case-Based Reasoning (CBR) technique, which has recently become widely used. Toward this end, an improved CBR model that uses the Multiple Regression Analysis (MRA) technique in the revision phase of the CBR technique was developed. To verify the prediction performance of the proposed model, a case study was performed on 41 business facilities and 99 multi-family housing projects. The results showed that the prediction performance of the revised CBR model for business facilities and multi-family housings improved by 17.23% and 4.39%, respectively, compared to that of the existing CBR model. The proposed MRA-based revised CBR model is expected to be useful in estimating the construction cost in the initial phase of a project.

1. Introduction

Accurately estimating the cost of a construction project is a task that is critical in the project's success. Therefore, a cost prediction model based on limited information in the initial phase of a construction project is essential (Hegazy, 2002).

Many researchers have studied cost prediction in the initial phase of a project using various techniques, including statistical-based analysis techniques (Lowe, Emsley, & Harding, 2006; Phaobunjong, 2002; Trost & Oberlender, 2003) such as Multiple Regression Analysis (MRA), probabilistic-based analysis techniques (Nassar, Gunnarsson, & Hegab, 2005) such as Monte Carlo Simulation (MCS), and artificial intelligence-based techniques (Attalla & Hegazy, 2003; Hegazy & Ayed, 1998; Kim, Seo, & Kang, 2005; Soutos & Lowe 2005) such as Artificial Neural Networks (ANN) and Genetic Algorithms (GA).

Recently, an increasing number of studies have used the Case-Based Reasoning (CBR) technique (An & Kang, 2005; Doğan, Arditi, & Gunaydin, 2006, 2008; Karshenas & Tse, 2002; Kim, An, & Kang, 2004; Kim & Kang, 2004; Koo, Hong, Hyun, & Koo, 2010; Marir & Watson, 1995; Perera & Watson, 1998; Yau & Yang, 1998). Most of them, however, have focused on the retrieval phase of the CBR cycle in terms of how the CBR techniques could be applied in predicting construction cost effectively, and only few of them have considered the methods for improving the CBR model's prediction performance.

Therefore, this study was aimed at improving the prediction performance of the CBR-based cost prediction model, targeting the widely used CBR technique. Toward this end, a revised CBR model was developed by using a MRA-based revision method to the revision phase of the CBR cycle.

2. Literature review

CBR is a data-mining technique that remembers similar situations applied to the solution of previous problems, and uses the information and knowledge from such situations to solve a new problem (Aamodt & Plaza, 1994). This technique does not require a clear model for problem-solving; rather, establishing cases is an important task in problem-solving. Since the model can be established by defining key attributes that express cases, this technique makes it easy to establish a model. It also allows simple maintenance and control of a vast amount of information by using a database technique (Kim & Kang, 2004). As shown in Fig. 1, the CBR technique has four phases (Aamodt & Plaza, 1994):

(i) Retrieve: This phase inquires about and extracts the previous case most similar to the current one from a case base;
(ii) Reuse: This phase reuses the information and knowledge from the retrieved case for problem-solving;

* Corresponding author. Tel.: +1 765 404 4954.
E-mail address: chokm76@nate.com (K. Cho).

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.11.018

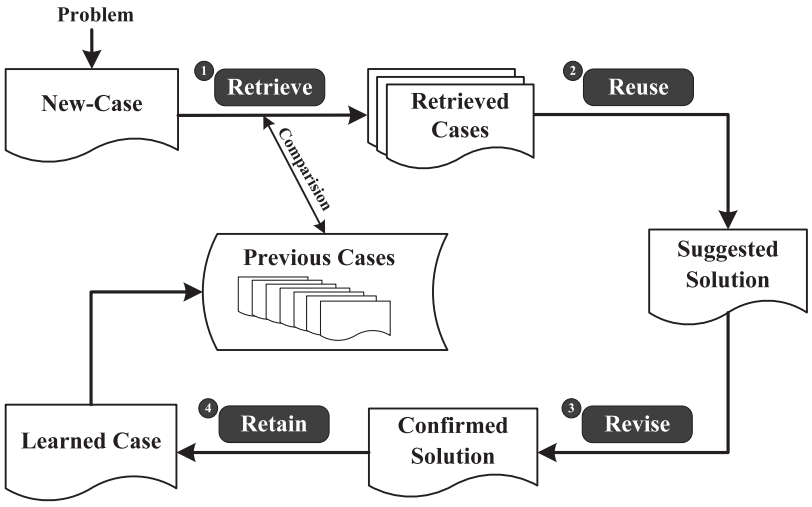

Fig. 1. The general cycle of CBR (Aamodt and Plaza 1994).

(iii) Revise: If the retrieved case is not suitable for solving the new problem, this phase analyzes the difference(s) between the new problem and the retrieved case, then revises the retrieved case accordingly; and
(iv) Retain: This phase stores in the case base the solution proposed from the retrieved case so that it can be used in future problems.

Various studies have used the CBR technique to predict construction cost in the initial phase of a project. Marir and Watson (1995) used CBR based on the recognition of the usefulness of previous experiences in predicting the construction cost. As a result, they confirmed that the CBR technique could be appropriately used in cost prediction. Yau and Yang (1998) demonstrated that the CBR technique could be used in construction management and in predicting the construction duration and cost in the preliminary design stage of a construction project. Kim, An, and Kang (2004) developed the cost prediction models based on CBR, MRA, and ANN, targeting a multi-family housing project, and compared the prediction performance of each models. The result showed that while the ANN model was somewhat better than the other models in terms of prediction performance, the CBR model was superior to the other two models in terms of long-term use, available information from the results, and time-versus-accuracy tradeoffs.

Since it was established that the CBR technique can be used to predict the cost of a construction project, studies have continued to improve the prediction performance of the CBR technique. An and Kang (2005), Doğan et al. (2006), Doğan et al. (2008), and Koo et al. (2010) focused on applying the optimal attribute weights in the retrieval phase of the CBR cycle to retrieve the case that is most similar to the test case. In doing so, they tried various methods, including Feature Counting (FC), Gradient Descent Method (GDM), Genetic Algorithm (GA), Analytic Hierarchy Process (AHP), Multiple Regression Analysis (MRA), and Decision Trees. This approach is limited, however, in that even if the most appropriate attribute weights are applied, it is difficult to present a solution (i.e., the construction cost) with a high level of accuracy if the degree of similarity of the retrieved case to the test case is low.

To overcome this limitation, recent studies have proposed the application of human intervention (Perea & Watson, 1998), adaptation of experienced estimators (Karshensa & Tse, 2002), and modification by experts (Yau & Yang, 1998; An & Kang, 2005; Koo et al., 2010) in the revision phase of the CBR cycle. However, revision by humans is a deterministic method as it depends on subjective decisions; and therefore, it lacks reliability, takes too much time to process, and is difficult to directly implement on a model.

Therefore, in this study, a cost prediction model was developed by applying the MRA method, a stochastic method based on real data, to the revision phase of the CBR cycle, in order to improve the prediction performance of the CBR-based cost prediction model. Shown in Table 1 are the revision phases and revision methods of previous studies that use the CBR technique.

3. Development of the MRA-based revised CBR model

As discussed in the Literature Review, existing cost prediction models using the CBR technique mostly focused on the retrieval phase. These models generally have three steps: (i) setting a case base; (ii) determining the attribute weights; and (iii) retrieving the similar case. However, the CBR model that is proposed in this study has four steps: the previous three steps and the revision step, as shown in Fig. 2. The detailed tasks for each step are as follows.

3.1. Step 1: Setting a case base

Developing a CBR-based cost prediction model requires the collection of existing cases and the establishment of a case base. Each case in the case base includes collectable information (i.e., attributes) in the step in which the model attempts to predict the cost. There are various methods to select attributes for retrieving similar cases. Kim and Kang (2004) used attributes from previous studies, An and Kang (2005) derived attributes from interviews with experts, and Koo et al. (2010) used those attributes with the highest frequency of use among the attributes used in previous studies.

In this study, the method of selecting the attributes consists of three steps. First is collecting the cases based on attributes that could be deliverable at the initial phase of a project. Second is conducting the MRA for screening the attributes that significantly affect the construction cost prediction. For example, if the significant p-value of the 'landscape area' is 0.15(>0.05) calculated through MRA, the 'landscape area' does not belong to the significant variables for predicting the construction cost with 95% significance level. In this study, therefore, these attributes were excluded in the CBR model. Third, the remaining attributes were used for the CBR model. As shown in Fig. 2, an MRA was conducted to derive the attributes.