Automation in Construction 96 (2018) 65–74

Contents lists available at ScienceDirect

Automation in Construction

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

Learning method for knowledge retention in CBR cost models

Sae-Hyun Ji[a], Joseph Ahn[b,*], Eul-Bum Lee[c], Yonggu Kim[c]

[a] Institute of Construction and Environmental Engineering, Seoul National University, Seoul 08826, Republic of Korea
[b] Division of Architecture, Civil and Environmental Engineering, Hoseo University, Asan 31499, Republic of Korea
[c] Graduate Institute of Ferrous Technology, POSTECH, Pohang 37673, Republic of Korea

ARTICLE INFO

Keywords:
Learning method
Knowledge retention
Case-based reasoning
Construction cost estimation

ABSTRACT

The case-based reasoning methodology fundamentally relies on historical cases to solve new problems. Supplementing insufficient data by the reproduction of appropriate values can mitigate the potential negative effects on the solutions resulting from sudden changes. However, CBR researchers have rarely examined this issue. To address this challenge, this research proposes a learning method for knowledge retention based on CBR by applying a data-mining approach to manage missing dataset values. A case study on a 164-apartment project was conducted to compare the estimation accuracy of the suggested learning method to that of past research with the same experiment conditions. The learning method with the CBR model achieved higher accuracy of the overall cost estimation and higher stability compared with the previous model. This research shows how cases can be generated and retained as learned cases to overcome the difficulties of continuous updates in a wide range of construction projects, as well as why the case bases need to be continuously updated. The research outcomes could support work related to cost estimation for decision makers ranging from beginners to experts in both academia and industry.

1. Introduction

An artificial intelligence (AI) program called "AlphaGo" has won consecutive matches of the game called "go," which is an example of the progress toward general-purpose AI applications. People are oblivious to the numerous AI applications that are based on machine-learning algorithms that exist in their lives. Likewise, in the construction domain, case-based reasoning (CBR) has gained much attention as a prominent machine-learning approach. This method has been applied in various ways, such as cost estimation, decision support, scheduling, safety, and litigation. These machine-learning systems are not sustainable without learning, which is the act of acquiring or modifying new or existing knowledge to lead to a potential change [13]. Thus, continuous learning stimuli and feedback are required. Stimuli and feedback are produced and retained in the form of data, and the general distribution characteristics should be considered to build a successful learning strategy.

In probability theory, the bell-shaped Gaussian distribution is very commonly used in the natural and social sciences to represent real-value random variables. As the name implies, the spread of the values is biased toward the center. Thus, there is a possibility of sudden feature shifts having a negative impact, which should be considered if the data at the ends of the distribution are used for problem solving. If insufficient data are supplemented by the reproduction of appropriate values, the potential negative effects of sudden changes on the solutions can be exaggerated, but this issue is rarely examined in most CBR research.

To deal with this challenge, this research proposes a learning method to ensure the knowledge retention in the CBR model. This is achieved by applying a data-mining approach for the management of missing dataset values. The learning shapes the biased bell shape of raw data into a quantitatively and uniformly expanded dataset. Training samples are generated based on the dependent value that is the most influential factor in the CBR model, and then the samples are temporarily reproduced as new cases.

The proposed method performance was evaluated by comparing the outcomes with those of a previous model, and then a surviving dataset was fed back into the model to update the algorithm. A case study was also conducted, and the estimation accuracy of the suggested method was compared with that of past research [18] using the same experimental conditions. As shown in Fig. 1, this study follows the research process established by Salkind [31]. This research begins with the question, "How can the performance of CBR be enhanced?"; accordingly, the authors' believe that it can be enhanced with the use of an effective learning method. In this regard, the key factors should be identified to develop the learning method of knowledge reproduction,

* Corresponding author.
E-mail addresses: oldclock@snu.ac.kr (S.-H. Ji), josephahn@hoseo.edu (J. Ahn), dreblee@postech.ac.kr (E.-B. Lee), ch61park@postech.ac.kr (Y. Kim).

https://doi.org/10.1016/j.autcon.2018.08.019
Received 8 December 2017; Received in revised form 21 June 2018; Accepted 29 August 2018
0926-5805/

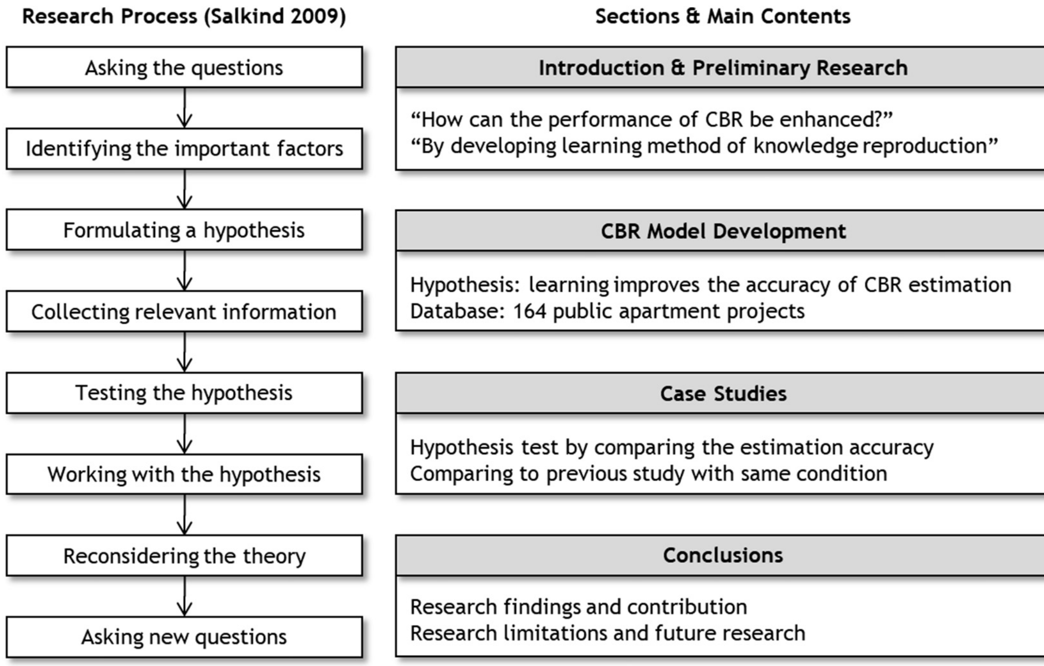

Fig. 1. Research process and methodology.

thereby leading to the formulation of the following hypothesis: "Learning improves the accuracy of the CBR cost estimation." To test this hypothesis, a CBR cost model was developed using a dataset of 164 public apartment projects, which is the same case examined in previous research [18], and the model was validated. The accuracy of the cost estimation was compared with that of the previous study with the same conditions. The knowledge gained from this research is discussed together with future research directions.

2. Preliminary research

2.1. Learning

Learning is the act of acquiring, modifying, or reinforcing new or existing knowledge, behaviors, skills, values, or preferences, which may lead to a potential change in the synthesis of information, the depth of knowledge, and the attitudes or behaviors that are related to the types and range of experience [13]. The ability to learn is possessed by humans, animals, plants [20], and some machines. This learning does not happen at once and involves a process built upon and shaped by previous knowledge. Psychologists usually define learning as "a relatively permanent change in behavior due to past experience; the process by which relatively permanent changes occur in behavioral potential as a result of experience [5]."

2.2. Case-based reasoning

The origin of CBR is cognitive science, which deals with psychological theories of the way that human memory works [26]. Schank [32] asserts that humans use cases to settle problems, and the process of remembering is a vital process of intelligence. People try to find matching cases and require more experience when inappropriate matches occur. The CBR methodology was developed based on primary work with human-memory. The basic idea is that similar problems have similar solutions. Thus, CBR involves storing a set of solution examples and adjusting previously used solutions to solve given problems [29]. Generally, the CBR problem-solving process consists of the following four steps: retrieval, reuse, revision, and retention [1]. CBR has been broadly applied across industries and used for discovering medical knowledge [10,11,27,39], managerial-decision support [2,33], healthcare management [16], educational applications [15], the diagnosis of power-transformer faults [28], cost estimation [4,6,19,21,22,37,38], international-market selection [25], decision-making support [7,8,24], planning/scheduling [21,30,34,37], safety-hazard identification [12], and the prediction of litigation outcomes [35].

At the conceptual level of the CBR model, appropriate cases are retrieved from a database to solve a problem and reused to treat the problem. If the suggested cases are not a close match, then the solutions need to be revised. The revised solutions are then retained as new cases in the database [36]. In an experience-oriented industry such as construction, the knowledge and assessment of previous projects are typically essential for the management of reoccurring problems. However, the time frame of the performance appraisal of a construction project is protracted because the appraisal cannot be finalized until the completion of the project. The case-base abundance in learning-based CBR results in time-consuming work, and this challenge has only been slightly addressed in previous research. Nevertheless, like other machine-learning algorithms, case-retention learning is an incremental process of the CBR algorithm.

2.3. Data mining

Data mining is frequently used as a preliminary process of discovering meaningful patterns and relationships within big data. It includes diverse types of processing that act on and prepare raw data for further processing [14]. The process starts with the conversion of the data into a structured format for easier and more effective processing that satisfies user-specific purposes. Several methods can be used to manage the missing values of a dataset [14]: 1) dismissal of the tuples, 2) manual input, 3) use of a global constant, 4) application of each attribute's average, 5) application of the attribute averages on the same class of a given tuple, and 6) producing the most probable values by an analysis such as regression or machine learning.

3. Learning algorithm

To prevent unexpected solution distortions due to sudden feature shifts, the case distribution should reflect the features of the solution