

Case-Based Reasoning Cost Estimation Model Using Two-Step Retrieval Method

Hyun-Soo Lee¹, Kihoon Seong², Moonseo Park³, Sae-Hyun Ji⁴, and Sooyoung Kim⁴

(Received August 31, 2010 / Revised October 5, 2010 / Accepted October 13, 2010)

ABSTRACT

Case-based reasoning (CBR) method can make estimators understand the estimation process more clearly. Thus, CBR is widely used as a methodology for cost estimation. In CBR, the quality of case retrieval affects the relevance of retrieved cases and hence the overall quality of the reminding capability of CBR system. Thus, it is essential to retrieve relevant past cases for establishing a robust CBR system. Case retrieval needs the following tasks to obtain appropriate case(s); indexing, search, and matching (Aamodt and Plaza 1994). However, the previous CBR researches mostly deal with matching process that has limits such as accuracy and efficiency of case retrieval. In order to address this issue, this research presents a CBR cost model for building projects that has two-step retrieval process: decision tree and nearest neighbor methods. Specifically, the proposed cost model has indexing, search and matching modules. Features in the model are divided into shape-based and scale-based attributes. Based on these, decision tree is established for facilitating the search task and nearest neighbor method was utilized for matching task. In regard to applying nearest neighbor method, attribute weights are assigned using GA optimization and similarity is calculated using the principle of distance measuring. Thereafter, the proposed CBR cost model is developed using 174 cases and validated using 12 test cases.

Key words Cost, Estimation, Case-based Reasoning, Genetic Algorithms, Decision Tree

1. Introduction

Case-based reasoning (CBR) is a problem solving approach that utilizes the specific knowledge of previously experienced and concrete problem situations to solve a new problem (Aamodt and Plaza 1994). The principle of CBR is similar to the point that estimate experts employ accumulated experiences and knowledge to assess cost of a new construction project. In this context, cost estimations using CBR are more understandable than other Artificial Intelligence (AI) approaches. Moreover, it is likely that estimators are willing to trust the solution regardless of speaker's reputation or background (Ji et al. 2010). Therefore, CBR is widely used as a methodology for cost estimation in construction domain. Generally, CBR has four problem solving processes; retrieve, reuse, revise and retain (Aamodt and Plaza 1994). One of the major issues lies in the retrieval of appropriate cases (Hansen 1995). Case retrieve is the process of searching and determining relevance or similarity of past cases in the case base with a new case. The objective of retrieve is to select out the most

useful previous cases towards the optimal resolution of a new case and to ignore those previous cases that are irrelevant (Ali and Kalyan 1997). As case retrieval is a primary process of CBR cycle, its result affects the following process such as reuse, revise and retain (Kolodner 1993). Furthermore, its quality directly affects the relevance of retrieved cases and the overall quality of the reminding capability of a CBR system (Goh and Chau 2009). As a result, it is essential to retrieve relevant previous cases for establishing a robust CBR system. More specifically, for conducting a retrieval process, a task should be decomposed to indexing, search, and matching tasks (Aamodt and Plaza 1994). Case retrieval algorithms can be sufficiently effective and efficient if the task decomposition is properly performed. However, previous CBR researches mostly deal with matching without considering the search task. Thus, it takes a long time to evaluate the relevance of cases because there are too many cases for the matching task. Furthermore, it may involve irrelevant cases with the new case. As a result, improper cases can be extracted for solving the given problem that has a negative impact

1) Prof., Dept. of Architecture, Seoul National Univ., San 56-1 Shinrim-dong, Seoul, Korea (Corresponding author: hyunslee@snu.ac.kr)

2) Research student, Dept. of Architecture, Seoul National Univ.

3) Associate Prof., Dept. of Architecture, Seoul National Univ.

4) Ph.D. candidate, Dept. of Architecture, Seoul National Univ.

5) Ph.D. student, Dept. of Architecture, Seoul National Univ.

on estimation accuracy and retrieval efficiency.

In order to address those issues, this research presents a CBR cost model that has two-step retrieval process for public apartment projects at conceptual stage. Especially, this research aims to propose a cost model focusing on case retrieval. The procedures of the research are as follows. First, the research scope is defined as an initial stage of construction projects. Subsequently, data is prepared for cost model development. A CBR cost model is developed mainly focusing on two step case retrieval approach which uses decision tree, and genetic algorithms method. Finally, the reliability of CBR cost model is validated by comparing with others types of cost model.

2. Previous Research

2.1 Case-Based Reasoning

CBR has its roots in psychological theory of human reasoning, which has the intuitive paradigm that humans solve new problems (Mount and Liao 2001). Instead of relying solely on general knowledge of a problem domain, or making associations along generalized relationships between problem descriptors and conclusions, CBR is able to utilize the specific knowledge of previously experienced, concrete problem situations. A new problem is solved by finding a similar past case, and reusing it in the new problem situation. A second important difference is that CBR also is an approach to incremental, sustained learning, since a new experience is retained each time a problem has been solved, making it immediately available for future problems (Aamodt and Plaza 1994). CBR contains two main assumptions; the first is that similar problems have similar solutions and the second is that once happened problem tends to come about again. In other words, CBR basically solves a new problem by remembering a previous similar situation and by reusing information and knowledge of past situation (Watson and Marir 1994). According to those characteristics, CBR has been used in the field of experience-oriented industry such as construction that knowledge and assessments of previous projects are essential for solving reoccurring problems (Ji et al. 2010). Many studies in construction domain related to CBR have been conducted cost estimation purposes (Karshenas and Tse 2002; Yi 2006; Chou 2009), international market selection (Ozorhan et al. 2006), construction claim (Arditi et al. 1999), scheduling (Ryu et al. 2007), and decision-making support (Chau et al. 2001).

2.2 Decision Tree

Decision tree is a tool for making number-based decisions where a large amount of complex information needs to be taken into account. It builds up a structure of tree using recursive partitioning. It provides an effective structure in which alternative decisions and the implications of taking those decisions can be

laid down and evaluated (Arditi and Pulket 2005). The decision tree is used for the purpose of data classification. Given a set of cases, a decision tree that classifies the set can be established. The decision tree is composed of node, link and leaf. Each node represents an attribute, and link corresponds to the possible value of the attribute. Each leaf shows a decision. The main issue is to decide on the root node. It is determined by how well each attribute classifies the case set. This research constructs the decision tree based on ID algorithm. It is widely adopted to generate a decision tree. The ID3 infers decision trees by growing them from the root downward, greedily selecting the next best attribute for each new decision branch of the tree. It prefers traversing as few nodes in the tree as possible (Xu et al. 2006).

3. Cost Model Development

3.1 Cost Model Scope

Conceptual cost estimation can be defined as the forecast of project costs that is performed before any significant amount of information is available from detailed design and with a still incomplete work scope definition (Bley 1990). It is used to tell the owner about the anticipated cost, thus presenting useful information for the owner in contemplating the project feasibility and further development (Barrie and Paulson 1992). Moreover, the output is the basis for the release of funds for further studies of estimates and become the marker against which all subsequent estimates are compared (Smith 1995). However, conceptual estimating is an inexact process based to a large degree on judgment and experience due to the lack of information and existing uncertainty at this stage (Bley 1990). Additionally, the time and cost available for making the estimate is restricted. Therefore, this research concentrates on the initial stage of construction projects. The object is a public housing construction projects.

3.2 Data Analysis

It is necessary to collect sufficient amount of data to develop the cost model using CBR. Primarily, the data is the most crucial input to cost model development since CBR utilizes the past cases. This research collects the data of 174 apartment buildings from 18 housing complex projects in Korea which covers from 2005 to 2008. All the historical data are supplied by a public enterprise by the Korean government. A construction project is typically a major and unique undertaking. Due to its uniqueness, it is extremely difficult to collect large amount of cost data for buildings relating to the same reference point, which is under the same time, location, and condition. In this manner, the collected cost data is composed of construction projects that are performed under different time and location. As a result, it is necessary to adjust each historical data into the same point or a common reference point. Thus, this research employed the Korean

government's historical cost index (KICT 2010). This process is referred to as the normalization of the cost data. Although the data should be normalized in terms of regional location, this research only makes the data normalized historically. Mainly due to Korea's relatively small territory, there is little necessity in normalizing the data for location (Ji et al 2010).

3.3 Cost Model Scheme

The cost model consists of three components; indexing module, search module and matching module. It is shown in Fig. 1. The indexing module identifies attributes and classifying them to reflect the characteristics of information. In search module, cost model aims at extracting case group that is potentially similar to new case. Finally, the best case(s) is selected by assessing case similarity in matching module. Case retrieval process, no matter the method, requires a combination of search and matching (Kolodner 1993). Thus, it is needed to take into account the means to combine them. Fig. 2 provides a structure of cost model. This research was supported by a grant (R&D06CIT-A03) from the Construction Technology Innovation Program funded by the Ministry of Land, Transport and Maritime Affairs of the Korean government.

Through decision tree, the input case conforms to one of 4 case groups. The only selected group becomes candidate for matching. Thus, on average, the number of case for matching declines to one forth.

3.4 Indexing Module

A case is composed of the problem features and their solutions. With these two case components, new problems can be solved in

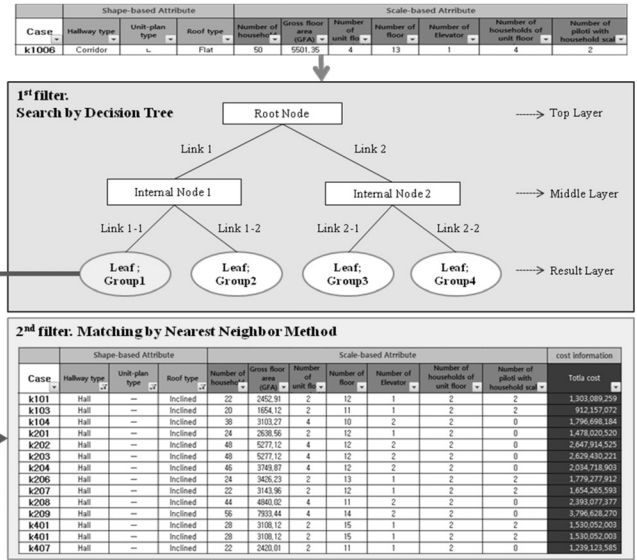


Fig. 2. Structure of Cost Model

CBR by first finding a relevant past case(s) and then adapting solution to that problem to the new situation (Kolodner 1993). To facilitate the case retrieval, it is required to assign indexes to cases. The main issue of case indexing is deciding problem features to ensure that case(s) can be retrieved at appropriate times. Accordingly, it is essential to identify problem features that can represent the case. Case retrieval process starts with identifying problem features (Watson and Marir 1994). As a descriptor of case, problem features explain the state that the case was happening. Problem features contain several types of information. The characteristics of information can be classified by the following criteria; the domain, the forms of representation, and the contents. For instance, some features have qualitative information; in contrast, others do quantitative information. Certainly, qualitative one is represented by textual form, and quantitative one is done by numerical form. Furthermore, the influence on case solution is determined by them. Consequently, it is required to classify the characteristics of information that problem features have. This research considers the content of information contained in problem features. The problem feature can be defined as an attribute in CBR. Based on the previous research and user interview, the pool of impact factor is mode. Then, we analyzed the sample of conceptual drawings comprehensively. This make the number of the factors decreased. Finally, those are confirmed by the expert group from 8 public enterprises. The attributes includes 10 of input attributes and an output attributes. This is shown in Table 1. The cost information of apartment building project is an output attribute. The extraction of input attributes is affected by the point that a cost model is developed for the initial stage as conceptual phase. The input attributes are divided into 2 groups; 3 of shape-based input attributes and 7 of scale-based input attribute.

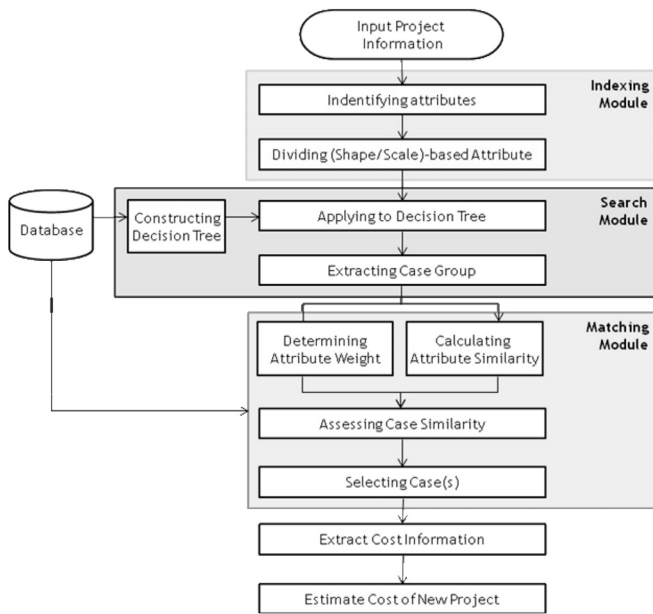


Fig. 1. Cost Model Scheme

Table 1. Classification of Attributes

Classification	Attribute		Range
Shape-basedinput attribute	X1	Unit-plan type	"┌" type or "—" type
	X2	Hallway type	Hall or Corridor
	X3	Roof type	Flat or Inclined
Scale-basedinput attribute	X4	Number of household	Integer
	X5	Number of floor	Integer
	X6	Gross floor area	Real number
	X7	Number of unit floor household	Integer
	X8	Number of piloti with household scale	Integer
	X9	Number of elevator	Integer
	X10	Number of households of unit floor per elevator	Real number
Output attribute	X11	Total Cost	Real number

3.5 Search Module

Case base stores a large number of records. CBR methodology is useful for large-scale problems only when its retrieval algorithms are sufficiently effective and efficient to handle a number of cases. The effectiveness of case retrieval is related to search task rather than matching task. The search algorithm is developed by using decision tree. During indexing module, this research extracted attributes, and divided them into shaped-based and scale-based attributes. Because shape-based attributes contains information of general domain, only shape-based attributes are considered in search algorithm. An important challenge is to decide the node. The following issues are required to be considered. The attribute should be determined for root node. This research defines the criteria for this problem. One is the ability of discrimination. Cases classified by a certain attribute has similar characteristics, so that they don't need to be discriminated. The subsequent one is influence to the output attribute. The more influential to cost a certain attribute is, the earlier it should come. Another critical thing to do is determining the number of layer. This research takes aim at distinguishing case group, so that cases of each group have similar characteristics in respect with shape-based attributes. However, decision tree should not result in excessive decrease of cases caused by several classifications. By considering the issues, this research constructed a decision tree based on defined criteria. This is represented in (Fig. 3).

A hallway type is selected for root node. This attribute is able to produce a leaf. The case group 4 is composed of data that feature corridor type for hallway. Most of them are also characterized by flat roof and "L" unit-plan type. Moreover, it has the most influential effect on project cost. Subsequently, the attribute of unit-plan type generates a leaf. Because cases in group 1 have the same roof type, it is unnecessary to classify them again. Finally, the remaining cases can be divided by roof type. Cases of flat roof type belong to group 2, and cases of inclined roof type do group 3. Consequently, the case data is categorized

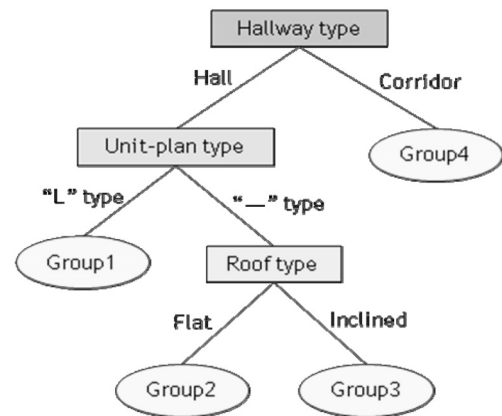


Fig. 3. Decision Tree for Search Task

into 4 case groups.

In case that roof type and unit-plan type is chosen as a root node, case group does not have similar characteristics. In addition, some case groups contain small number of cases, so that the cost model shows low degree of performance.

3.6 Matching Module

This research adopts a nearest neighbor method for matching task. This involves the assessment of similarity between stored cases and the new input case, based on matching a weighted sum of features. There are challenges related to the nearest neighbor method. The first thing is to assign the attribute weight values. The second challenge is the computation of similarity. And, because scale-based attributes contains information of specific domain, only scale-based attributes are considered in matching task.

3.6.1 Assigning Attribute Weight Value

The weight assigns a value of importance to each attribute. In general, retrieval of the most relevant case is determined by the

presence of a greater number of higher priority attributes matching between the new case and the stored cases (Dogan et al. 2008).

In this research, the attribute weight value is optimized using GAs. GAs are used to search a space of candidate solution to identify the best one. The best hypothesis is defined as the one that is the optimized value to the pre-defined numerical measure at hand, which is called hypothesis fitness (Mitchell 1997). To facilitate the research model we adapt an algorithm proposed by Ji et al. (2010) shown in Equation 1.

$$\text{Minimize } \sum_{k=1}^n |D_k| \quad \text{s.t. } \begin{pmatrix} C_1 \\ \vdots \\ C_j \end{pmatrix} - \begin{pmatrix} X_{i1} & \dots & X_{ij} \\ \vdots & \ddots & \vdots \\ X_{i1} & \dots & X_{ij} \end{pmatrix} \times \begin{pmatrix} W_1 \\ \vdots \\ W_j \end{pmatrix} = \begin{pmatrix} D_1 \\ \vdots \\ D_j \end{pmatrix} \quad (1)$$

Where, C_j denotes project cost for case j , X_{ij} denotes j th attribute value for case i , W_j denotes weight for j th attribute, and D_i denotes distance for case i .

3.6.2 Calculating Attribute Similarity and Case Similarity

The computation of similarity is an important issue for the case retrieval in CBR. An appropriate similarity function needs to be developed to handle the hidden relationships between the objects associated with cases (Burkhard 2001). This research computes the attribute similarity by employing the principle of distance measuring. As scale-based attributes are the types of quantitative data, the similarity can be calculated by the numerical function as Equation 2.

$$AS_i = 1 - \frac{\sqrt{(X_{Ni} - X_{Ci})^2}}{\sqrt{(X_{Hi} - X_{Li})^2}} \quad (2)$$

Where, AS_i denotes similarity score for attribute i , X_{Ni}, X_{Ci}, X_{Hi} , and X_{Li} denotes new case value, compared past case value, highest past case value and lowest past case value for attribute i .

Measuring the case similarity is the last step for nearest neighbor method. According to the definition of nearest neighbor, case similarity can be measured by Equation 3.

$$f_{CS}(x) = \frac{\sum_{i=1}^n (f_{AS_i} \times f_{AW_i})}{\sum_{i=1}^n (f_{AW_i})} \quad (3)$$

Where, n denotes the number of attribute, and f_{CS}, f_{AS} , and f_{AW} denotes function of case similarity f_{AW} , attribute similarity, and attribute weight.

4. Case Study

4.1 Outline of Case Study

To validate the accuracy of model, it is required to choose test case(s). The random selection of the test case may affect the accuracy of testing. In other words, the test results are likely to change when different testing cases are selected (Dogan et al. 2008). Therefore, this research chooses 12 test cases in case base. The selected case data set is utilized only for case study, and excluded from cost model development. The profile of 12 cases is shown in table 2. Then, the accuracy of the proposed cost model is measured by error rate. The error rate is calculated in absolute value. Finally, since the major difference of the proposed cost model is case retrieval approach, the accuracy is compared with other case retrieval method. Except for search and matching, identical processes of selecting cases and reusing the cost information are applied to the compared case retrieval method.

4.2 Results and analysis

The accuracy of cost model is calculated by the following

Table 2. Profile of Test Case

Case number	Shape-based Attribute			Scale-based Attribute							cost information
	Unit-plan Type	Hallway Type	Roof Type	Number of house hold	Number of floor	GFA	Number of unit floor house hold	Number of piloti with house hold scale	Number of Elevator	Number of Households of unit floor per elevator	Total cost
1	—	Hall	Inclined	54	14	5932.77	4	0	2	2	2,905,716,406
2	—	Hall	Inclined	44	12	3367.97	4	4	2	2	1,735,231,945
3	—	Corridor	Inclined	56	14	4350.92	4	0	1	4	2,048,999,772
4	—	Hall	Flat	22	14	2541.74	2	0	1	2	1,148,868,556
5	└	Hall	Flat	50	14	7199.42	4	2	2	2	3,303,973,323
6	—	Hall	Inclined	19	10	1597.1	2	0	1	2	746,039,456
7	—	Corridor	Inclined	44	11	3154.68	4	0	1	4	1,473,618,278
8	└	Hall	Inclined	40	15	5642.07	3	2	1	3	2,735,539,839
9	—	Corridor	Inclined	52	13	3955.76	4	0	1	4	1,847,508,256
10	—	Hall	Flat	16	9	1728.37	2	2	1	2	831,014,480
11	—	Hall	Flat	44	12	3489.86	4	0	2	2	1,638,464,463
12	└	Hall	Flat	34	12	3660.87	3	2	1	3	1,887,097,156

process. First of all, it is necessary to identify which case group corresponds to each test case. All attributes are divided into shape-based and scale-based in the indexing part. The case group of test cases can be decided by applying shape-based attributes to decision tree. Table 3 represents the shape-based attributes and determined case group.

Subsequently, it is needed to assign attribute weight value. Attribute weight is decided by the case stored in case base. Because each case group contains different cases, it has different value of attribute weight. The process that GAs searches the optimized value is repeats for all case groups. The result is shown in table 4.

Then, case similarity can be calculated by adopting the concept of distance measuring. This assesses the similarity between the test case and cases stored in case base. Based on the attribute weight value and attribute similarity score, case similarity is computed. Lastly, this research retrieved the case(s) that case similarity scored over 0.9. Cost model utilizes the information of extracted case(s). The cost information denotes a total cost of building construction project. However, the unit cost (won/m²) is used as the output of proposed cost model. To estimate the cost of test case, the average unit cost of retrieved case(s) is multiplied by gross floor area of the test case. Table 5 presents the result of validation. It shows the actual cost, estimated cost and error rate. Consequently, the proposed CBR cost model results in 3.66% of mean error rate. AACE proposes the five classes of cost estimate based on level of project definition. And, it defines the expected

Table 3. Test Case Classification

Case number	Shape-based Attribute			Classified Case Group
	HallwayType	Unit-planType	RoofType	
1	Hall	—	Inclined	group 3
2	Hall	—	Inclined	group 3
3	Corridor	—	Inclined	group 4
4	Hall	—	Flat	group 2
5	Hall	⌊	Flat	group 1
6	Hall	—	Inclined	group 3
7	Corridor	—	Inclined	group 4
8	Hall	⌊	Inclined	group 1
9	Corridor	—	Inclined	group 4
10	Hall	—	Flat	group 2
11	Hall	—	Flat	group 2
12	Hall	⌊	Flat	group 1

Table 4. Result of Assigning Attribute Weight

Group number	Attribute Weight, W _j						
	W ₁	W ₂	W ₃	W ₄	W ₅	W ₆	W ₇
Group1	0.165	0.679	0.000	0.089	0.000	0.022	0.045
Group2	0.076	0.734	0.000	0.058	0.015	0.000	0.118
Group3	0.053	0.782	0.000	0.032	0.097	0.000	0.036
Group4	0.000	0.814	0.000	0.034	0.010	0.000	0.142

Table 5. Results of Implementation

Case number.	A Proposed Case Retrieval Approach (Search and Matching)		A Compared Case Retrieval Approach (Matching)	
	Estimated Cost (won)	Error rate	Estimated Cost (won)	Error rate
1	2,952,296,277	1.60%	2,992,622,278	2.99%
2	1,785,431,105	2.89%	1,818,411,745	4.79%
3	2,115,482,168	3.24%	2,008,020,402	2.00%
4	1,121,409,044	2.39%	1,222,971,122	6.45%
5	3,391,343,546	2.64%	2,896,846,325	12.32%
6	808,269,724	8.34%	778,505,120	4.35%
7	1,465,218,439	0.57%	1,422,927,699	3.44%
8	2,731,771,448	0.14%	2,764,702,743	1.07%
9	1,881,613,889	1.85%	1,842,844,516	0.25%
10	781,224,651	5.99%	841,360,884	1.25%
11	1,508,639,931	7.92%	1,724,654,250	5.26%
12	1,766,587,315	6.39%	1,759,925,311	6.74%
Mean		3.66%		4.24%
S.D		2.79%		3.30%

accuracy range for each class. When level of project definition corresponds to 10% to 40%, the accuracy is expected to be -10% to -20% and +10% to +30% (AACE 1997). The proposed cost model has error rate of 0.14% to 8.34% and mean error rate of 3.66%. This error rate is superior to accuracy rate proposed by AACE. Therefore, it is considered to be acceptable for conceptual cost estimation.

In addition, we verify the effectiveness of search algorithm, a case retrieval method without considering search task is examined. This involves the same attributes with the proposed approach. However, it applies 10 of input attributes to only matching task. Nearest neighbor method is adopted as matching algorithm. The identical process of assigning attribute weight value is applied for compared retrieval method. The result of comparison is provided in table 5. The compared retrieval method has error rate of 0.25% to 12.32% and mean error rate of 4.24%. In spite of small difference in average, this method presents also shows acceptable estimating performance. However, it has 3.30% of standard deviation. In compared with 3.30%, the proposed retrieval approach yields 2.79% of standard deviation. As a result, the suggested approach in this research has more stable performance than compared one.

5. Conclusion

Mainly due to the importance of conceptual cost estimation, this research aims at developing the CBR cost model using two-step retrieval process. The importance of this research is the proposal of case retrieval approach dealing with search and

matching tasks. Furthermore, the contributions of this research can be grouped into three parts. First, this research divided attributes based on contents in order to reflect the characteristics of information. Thus, shape-based attributes are used for search task, and scale-based attributes are utilized for matching task. Depending on the characteristics, it is required to differentiate the application of attributes. By grouping attributes that has similar characteristics, cost model using case-based reasoning is able to measure the influence of attributes on cost information. Second, this research established search algorithms by using the decision tree. By filtering, search algorithm eliminates inappropriate past cases from data set for matching. Without considering search task, cost model needs to evaluate a number of cases for matching, which takes a long time to check relevance of them. Moreover, it involves irrelevant cases with the new case in data set for matching. Therefore, it can improve the accuracy and efficiency of case retrieval. Finally, this research adopted GAs as a technique for assigning attribute weight. GAs are used to search a space of candidate solution to identify the best one. It has the strengths to overcome impasses that occur at local optima, and be more objective than AHP. However, it should be noted that this research is conducted with limited case data such as apartment building project in Korea. Thus, additional future studies are required that related to the number of k is required to be determined for k -nearest neighbor method; and application to other type of construction projects

Acknowledgment

This research was supported by a grant (R&D06CIT-A03) from the Construction Technology Innovation Program funded by the Ministry of Land, Transport and Maritime Affairs of the Korean government.

References

1. Aamodt, A. & Plaza, E. (1994). Case-Based Reasoning: Foundational Issues”, Methodological Variations, and System Approaches, *AI Communications*. IOS Press, Vol.7:1, pp.39-59.
2. An, S-H. & Kang, K-I. (2005), “A Study on Predicting Construction Cost of Apartment Housing Using Experts’ Knowledge at the Early Stage of Projects”, *Journal of Architectural Institute of Korea*, Vol.21, No.6, pp.81-88.
3. Arditi, D. & Tokdemir, O. B. (1999). “Using Case-based Reasoning to Predict the Outcome of Construction Litigation”, *Computer-Aided Civil and Infrastructure Engineering*, 14, pp.385-393.
4. Burkhard, H-D. (2001). *Similarity and Distance in Case-based Reasoning*, *fundamenta Informaticea*, 47, pp.201-215.
5. Dogan, S. Z. et al. (2006). “Determining Attribute Weights in a CBR Model for Early Cost Prediction of Structural Systems”, *Journal of Construction Engineering and Management*, Vol.132, No.10, 2006, pp.1092-1098.
6. Dogan, S. Z. et al. (2008). “Using Decision Trees for Determining Attribute Weights in a Case-Based Model of Early Cost Prediction”, *Journal of Construction Engineering and Management*, Vol.134, No.2, 2008, pp.146-152.
7. Duverile, P. & Castelain, J. M. (1999). “Cost Estimation During Design Step: Parametric Method versus Case Based Reasoning Method”, *Advanced Manufacturing Technology*, 15, pp.895-906.
8. Goh, Y. M. & Chua, D. K. H. (2009). “Case-Based Reasoning for Construction Hazard Identification: Case Representation and Retrieval”, *Journal of Construction Engineering and Management*, Vol.135 No.11, 2009, pp.1181-1189.
9. Goldberg, D. E. (2006). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley.
10. Ji, S-H. et al. (2009). “Cost Estimation Model for Building Projects Using Case-Based Reasoning”, submitted to the *Canadian Journal of Civil Engineering*, National Research Council of Canada.
11. Karshenas, S. & Tse, J. (2002). “A Case-Based Reasoning Approach to Construction Cost Estimating”, *Information Technology 2002, Computing in Civil Engineering*, pp.113-123.
12. Kim, G-H. & Kang, K-I. (2004), “A Study on Predicting Construction Cost of Apartment Housing Projects Based on Case Based Reasoning Technique at the Early Project Stage”, *Journal of Architectural Institute of Korea*, Vol.20, No.5, pp.83-92.
13. Kim, G-H. et al. (2004), “Comparing Accuracy of Prediction Cost Estimation Using Case-Based Reasoning and Neural Networks”, *Journal of Architectural Institute of Korea*, Vol.20, No.5, pp.93-102.
14. Kim, K. J. & Kim, K. (2010), “Preliminary Cost Estimation Model Using Case-Based Reasoning and Genetic Algorithms”, *Journal of Computing in Civil Engineering*, accepted January 21, 2010.
15. Kolodner, J. (1993). *Case-Based Reasoning*, Morgan Kaufmann Publishers, Inc., pp.289-389.
16. Mitchell, T. M. (1997). *Machine Learning*, McGraw Hill.
17. Montazemi, A. R. & Gupta, K. M. (1997). “A framework for retrieval in case-based reasoning systems”, *Annals of Operations Research*, 72, pp.51-73.
18. Phaobunbong, K. (2002). *Parametric Cost Estimating Model for Conceptual Cost Estimating of Building Construction Projects*, Doctor Dissertation, University of Texas at Austin, pp.10-50.
19. Ryu, H-K. et al. (2007). “Construction Planning Method Using Case-Based Reasoning (CONPLA-CBR)”, *Journal of Computing in Civil Engineering*. 21(6), ASCE, pp.410-422.
20. Seeley, I.H. (1997). *Quantity Surveying Practice 2nd Ed.*, Macmillan Press.
21. Shin, K-S. & Han, I. (2001). “A case-based approach using inductive indexing for corporate bond rating”, *Decision Support Systems* 32, pp.41-52.
22. Song, H. C. et al. (2007). “Case-Based Reasoning Method Using Case Data Base of Tall Buildings in Korea”, *Journal of the Korean Association for Spatial Structures*, Vol.7, No.6, pp.75-82.
23. Watson, I. & Marir, F. (1994). “Case-Based Reasoning: A Review”, *The Knowledge Engineering Review*, Vol.9, No.4, pp.6-13.
24. Yau, N-J. & Yang, J-B. (1998). “Case-Based Reasoning in Construction Management”, *Computer-Aided Civil and Infrastructure Engineering*, 13, pp.143-150.