S12-4

A STUDY ON THE DEVELOPMENT OF A COST MODEL BASED ON THE OWNER'S DECISION MAKING AT THE EARLY STAGES OF A CONSTRUCTION PROJECT

Choong-Wan Koo[1], Sang H.Park[2], Joon-oh Seo[3], TaeHoon Hong[4], and ChangTaek Hyun[5]

[1] Associate Researcher, HanmiParsons Co., Ltd., Seoul, Korea
[2] Corresponding Author. Senior Researcher, HanmiParsons Co., Ltd., Seoul, Korea
[3] Associate Researcher, HanmiParsons Co., Ltd., Seoul, Korea
[4] Assistant Professor, Department of Architectural Engineering, Yonsei University, Seoul, Korea
[5] Professor, Department of Architectural Engineering, University of Seoul, Seoul, Korea
Correspond to parksh@hanmiparsons.com

ABSTRACT: Decision making at the early stages of a construction project has a significant impact on the project, and various scenarios created based on the owner's requirements should be considered for the decision making. At the early stages of a construction project, the information regarding the project is usually limited and uncertain. As such, it is difficult to plan and manage the project (especially cost planning). Thus, in this study, a cost model that could be varied according to the owner's requirements was developed. The cost model that was developed in this study is based on the case-based reasoning (CBR) methodology. The model suggests cost estimation with the most similar historical case as a basis for the estimation. In this study, the optimization process was also conducted, using genetic algorithms that reflect the changes in the number of project characteristics and in the database in the model according to the owner's decision making. Two optimization parameters were established: (1) the minimum criteria for scoring attribute similarity (MCAS); and (2) the range of attribute weights (RAW). The cost model proposed in this study can help building owners and managers estimate the project budget at the business planning stage.

Keywords: *Case-based reasoning, cost planning, optimization*

1. INTRODUCTION

1.1 Background and Purpose

The construction industry has features that are in stark contrast to those of the manufacturing industry, which produce final products based on an order with a certain design in a particular site. The stakeholders in charge of a project are organized based on a particular project and are selected via bidding. It makes the construction industry distinctive. Since recently, as construction projects have become highly complicated, diversified, and bigger, the level of uncertainty of the success or failure is rising.

Decision making at the early stages of a construction project has a great effect on the project. With a project going forward, the specific information regarding it increases, which makes decision making more accurate. The time and efforts involved in the project also increase, however, and the level of effectiveness goes down.

Especially in the public sector, the industry often fails to break away from passive methods in which it barely manages to meet the budget presented by the policy. To overcome such a custom and to improve the competitiveness of the construction industry, more accurate information regarding critical factors, such as the construction cost, must be ensured at the early stages of a construction project (Koo et al. 2008; Koo 2007).

This study was conducted to improve the effectiveness of a construction project in the public sector. The model that was developed in this study requires the construction manager to engage in cost planning, depending on the owner's decision making at the early stages. This model was designed to coincide with the current practical process, to reflect a future change in the construction environment, and to suggest trusted performance.

1.2 Scope and Methodology

The cost model that was developed in this study was designed to be used at the early stages of a construction project. The cost data of public offices, such as municipal, district, and post offices, were used in this study. The model was divided into three parts: Architecture_Structure, Architecture_Finishing, and Others (landscape architecture, earthwork, mechanical work, electrical work, and communication work). The project information defined at the early stages of a project are very restrictive, but some information that could be analogized or assumed were used to develop the model.

One or more similar projects chosen from among the completed or ongoing projects are used as references in the practical budgeting process. The cost per square meter of these selected projects is applied to a new project. Case-based reasoning (CBR) and genetic algorithms (GA) were used to develop the model that was proposed in this study. These methodologies have a number of beneficial features, such as that it is not only most similar to the practical process but is also flexible and can thus reflect the changes in the business environment.

CBR is a method in which the most similar cases selected from among the historical data are applied to a new project. GA, on the other hand, is a method that can optimize the model in the event that certain project information or cases in the databases are changed. As shown in the previous researches using CBR, some factors with regard to attribute similarity should be stipulated, and some factors regarding the attribute weight were not easy to confirm in the CBR algorithm. To solve these problems, the optimization process was applied to this model, using GA. In the GA, some variables that have an effect on the target variable (i.e., prediction accuracy) were established as optimization parameters. In the optimization process, the GA finds the optimization value of these parameters within certain ranges.

The research process was as follows:

(1) The practical estimation process was figured out through an interview with the managers in charge of estimating the project budget, and the project information that have an effect on the decision making at the early stages of the project were analyzed through the interview.

(2) CBR, which is most similar to the practical process, was used to develop the model, and GA was applied to optimize some parameters that make CBR more efficient. The model was developed focusing on both the usability of the end user and the extendability of the model.

(3) A sensitivity analysis of the optimization parameters was conducted to determine the prediction capacity according to the change in the parameter value.

(4) As mentioned above, the proposed model was developed to improve the prediction capacity of the proposed model, where CBR and GA were applied. To validate the capacity of the model, the validation process was carried out by case application.

2. LITERATURE REVIEW

2.1 CBR Methodology

CBR is suitable for the most similar cases selected from among the historical data, which can be used as useful references. The results that will be obtained from the historical data can be presented as supporting evidences rather than as precise or accurate data.

As shown in Fig. 1, all the CBR methods employ the following 4RE process:

- REtrieve: During retrieval, the most similar cases are selected based on the retrieval parameters, through a comparison with the historical databases.
- REuse: During reuse, the case is adapted to fit to the current situation, to address the problem.
- REvise: The proposed solution is determined with some degree of uncertainty. If necessary, it is revised.
- REtain: During retention, the case is stored in case base, with an indicator of whether it was successful or not.

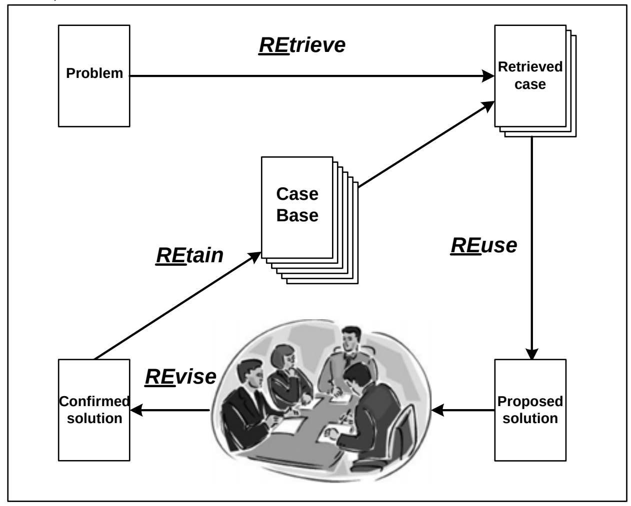

Figure 1. 4RE process of CBR

The CBR method is used for classification and synthesis tasks. Most of the CBR tool support classification tasks are related to case retrieval. On the other hand, synthesis tasks are used to find a new solution in addition to the existing solution. CBR is being applied in various fields, as shown in Table 1:

Table 1. CBR Application Fields and Specific Example

Class	Field	Specific Example
Classification tasks	Diagnosis	Medical diagnosis, machine defect diagnosis
	Prediction	Machine defect prediction, stock market prediction
	Assess-ment	Risk analysis of a bank or insurance, project cost assessment
	Process control	Process control related to machine equipment
	Planning	Travel plan, reuse of job schedule
Synthesis tasks	Design	Creation of a new design in addition to the existing design
	Planning	Creation of a new plan in addition to the existing plan
	Con-figuration	Creation of a new schedule in addition to the existing schedule

2.2 GA Methodology

GA is an adaptive heuristic algorithm based on the evolutionary concept of natural selection. It is designed to simulate the process of natural selection first identified by Charles Darwin in his “survival of the fittest” theory. As in this theory, GA introduces an intelligent algorithm that is a random search within a defined range to address a problem.

GA can provide benefits to anyone who wants to discover the best solution for difficult high-dimensional problems. Its performance is superior to those of other methodologies. The advantages of GA are its simplicity and speed as a search algorithm as well as its ability to discover solutions for the complicated problems. GA is useful and efficient when:

- the search range for a solution is large, complex, or poorly understood;
- the search criteria for a solution is very complicated, high-dimensional, or poorly understood;
- mathematical analysis cannot be applied; and
- the traditional search methods fail.

The GA approach can pursue complicated objectives with ease. All the objectives can be handled as weighted components of the fitness function, making it easy to adapt the GA scheduler or estimator to the particular requirements of a very wide range of possible overall objectives.

2.3 Comparison of Several Methods

The previous researches applied various methods to address the construction-related problems and to improve the accuracy of cost planning. Some of the methods that were used in the previous studies are as follows:

- analogical methods such as CBR (Koo et al. 2008; Koo 2007; Dogan 2006; Duverlie 1999);
- statistical methods such as multiple regression analysis (MRA) (Koo et al. 2008; Koo 2007; Phaobunjong 2002);
- repetitive learning methods such as the artificial neural network (ANN) (Koo et al. 2008; Koo 2007; Dogan 2006; Hegazy 1998); and
- optimization methods such as GA (Koo et al. 2008; Koo 2007; Dogan 2006).

It was found that the aforementioned methodologies should be applied to the proper fields according to the objective of using methodologies or distinct characteristics, such as the applied fields, data, and optimization level. CBR has characteristics that are similar to humans' heuristic approach, in which decisions are based on experience. GA has an algorithm that deduces the optimized value in the repeated and complicated process.

A model that integrates the advantages of CBR and GA has been studied (Koo et al. 2008; Koo 2007; Dogan 2006).

Koo et al. (2008) and Koo (2007) studied whether the CBR-based hybrid model employs the optimization process using GA, where the target is based on the prediction accuracy, which is different from the previous study (Dogan 2006), where the target was based on case similarity.

The results of the aforementioned studies proved that the CBR model that is integrated with GA not only has improved prediction accuracy but is also easy to optimize whenever the cost data are changed or whenever new cost data are added.

Moreover, in the study conducted by Koo et al. (2008) and Koo (2007), ANN, MRA, and MCS were combined besides CBR and GA, which were focused on prediction accuracy rather than on usability or simplicity.

3. THE CURRENT STATE OF COST PLANNING

The current state of cost planning (i.e., process, stakeholders, and services) was identified through extensive literature review and interviews with experts in the field of estimation. Interviews were conducted with public institutions like the National Police Agency, the National Statistical Office, the Supreme Court, and the Small and Medium Business Administration.

3.1 Approval Process for Public Offices

To obtain approval for a construction project from a public office, several organizations, such as those engaged in deliberation, admission, and demand, participate in the approval process. For example, in the case of the construction of a municipal office, the district ministry submits a report on the demand for a new building to the central ministry, which reviews the report and decides if a new building is indeed needed. After doing so, the central ministry devises a management plan for the supply and demand program of the public office. This plan is submitted to the Ministry of Public Administration and Security if the ministry approves the plan. The central ministry then submits a plan regarding the size of the office and the budget to the Ministry of Strategy and Finance. If the ministry approves the plan, the district ministry decides on the project delivery method and prepares the Request for Proposals (RFP). Below is a diagram of the aforementioned procedure.

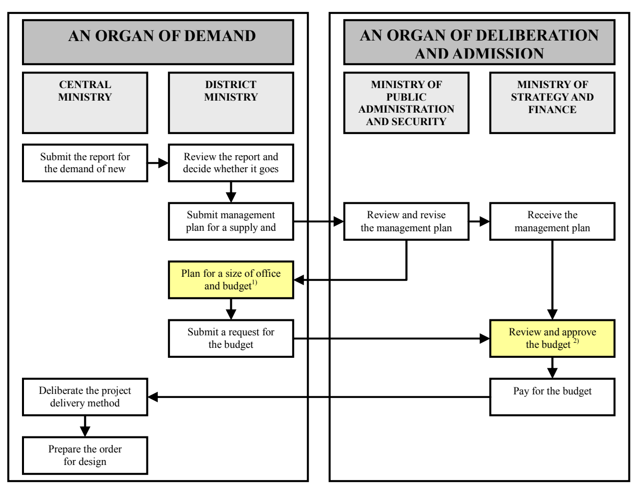

Figure 2. Approval process for the construction of public office

As shown in Fig. 2, there are two steps in cost planning. First, the central ministry, as an organ of demand, plans the size of the office and the project budget. Second, the Ministry of Strategy and Finance, as an organ of both deliberation and admission, reviews the budget and approves the plan.

Table 2 gives a detailed description of the aforementioned two-step procedure. First, in the step involving planning the size of the office and the budget, the most similar project would be selected from among the historical data. There is currently no systematic format, however, for keeping the cost data in good order. Second, in the step involving the review of the budget and the approval of the plan, since the review process depends on the subjective point of view of the man in charge of both deliberation and admission, the process lacks objectivity.

Table 2. Stakeholders and Services in relation to Cost Planning

Categories	Stakeholders	Services Assigned	Existing Problems
Plan for the size of the office and budget[1]	▪ The man in charge of finance in the central ministry as an organ of demand	▪ Plan regarding the size of the public office ▪ Cost planning using historical data	▪ Absence of a systematic format for keeping the data in good order ▪ Dependence on the data made by the supply office
Review and approval of the budget and plan[2]	▪ The man in charge of budget in the Ministry of Strategy and Finance as an organ of deliberation and admission	▪ Review and revision based on the budget submitted by the organ of demand ▪ Final approval of the budget and payment	▪ Lack of objectivity due to the dependence on the subjective point of view of the man in charge of deliberation and admission

3.2 Influencing Factors by Class

Table 3 presents the factors by class, which has a direct or indirect effect on cost at the early stage. The compulsory factors include the facility function, site location, plottage, total floor area, land ratio, floor space index, landscape area, public open space, no. of parking lots, no. of stories below the ground, and no. of stories above the ground, which would already be decided upon at the early stages of the project. The optional factors include the type of structure, the type of window, the external materials, the environmental grade, the communication grade in a finishing class, and the type of structure, environmental grade, and communication grade in a class of others, which would not be decided yet but could be analogized or assumed at this stage.

Table 3. Influence Factors by Class

No.	Influence Factor	Class		
		Structure	Finishing	Others
1	Facility function	●	●	●
2	Site location	●	●	●
3	Plottage	●	●	●
4	Total floor area	●	●	●
5	Land ratio	●	●	●
6	Floor space index	●	●	●
7	Landscape area	●	●	●
8	Public open space	●	●	●
9	No. of parking lot	●	●	●
10	Environment grade	-	O	O
11	Type of structure	O	O	O
12	No. of stories below the ground	●	●	●
13	No. of stories above the ground	●	●	●
14	Type of window	-	O	-
15	External materials	-	O	-
16	Grade on communication	-	O	O

● : compulsory factor, O : optional factor

4. MODEL DEVELOPMENT

It is assumed in this study that the cost model that integrates GA with CBR, which is focused on usability and simplicity, would be as accurate as the other cost estimating methods.

As presented in Table 3, there were optional factors as well as compulsory factors. Model I by class was developed only with compulsory factors, and model II was developed with optional factors in addition to compulsory factors. Therefore, six models were developed in this study. The other details are as follows.

4.1 Application of CBR

It is critical to calculate the attribute similarity and attribute weight in a CBR model. As the value of these parameters may be changed, the prediction accuracy could be very different. The nearest-neighbor retrieval method was used to calculate the attribute similarity, and GA was applied to calculate the attribute.

▪ Calculation of Attribute Similarity

For the attributes in the nominal scale, when the value of the attribute was the same, it was rated as 1; otherwise, 0. If an attribute was either in the interval or the ratio scale, it was scored based on equation [1] only when the score of attribute similarity was more than that of MCAS.

$$f_{AS}(x)=\begin{cases} 100-\left(\dfrac{\left|AV_{Test\_Case}-AV_{Retrieved\_Case}\right|}{AV_{Test\_Case}}\times 100\right) & if,\ f_{AS}(x)\geq MCAS \\ 0 & if,\ f_{AS}(x)<MCAS \end{cases} \quad [1]$$

where, f_{AS} is a function of attribute similarity, AV_{Test_Case} is the attribute value of the test case, $AV_{Retrieved\_Case}$ is the attribute value of the retrieved case, MCAS is the minimum criterion for scoring the attribute similarity.

▪ Calculation of Attribute Weight

In this study, the following two methodologies were used to calculate the attribute weight:

(1) Feature counting: This method applies 1 as a weight to all the attributes, based on the understanding that there is no need to apply to them a weight higher than 1. FC was the control group compared to GA.

(2) GA: This method optimizes the value of the attribute weight with the target based on the prediction accuracy, where the attribute weights could be changed within a range using GA.

▪ Calculation of Case Similarity

The method of calculating the attribute weight was introduced above. Equation [1] shows the method of calculating the attribute similarity. By multiplying these two values, the weighted-attribute similarity can be calculated. The accumulated sum of such value by attribute (attribute weight × attribute similarity) is divided by the accumulated sum of the attribute weight to calculate the case similarity score. The case similarity score was calculated using equation [2].

$$f_{CS}(x) = \frac{\sum_{i=1}^{n}\left(f_{AS_i} \times f_{AW_i}\right)}{\sum_{i=1}^{n}\left(f_{AW_i}\right)}, \left(n = the\ Number\ of\ Attributes \right) \quad [2]$$

where, f_{CS} is a function of case similarity, f_{AS} is a function of attribute similarity, f_{AW} is a function of attribute weight.

▪ **Analysis of Prediction Accuracy**

This study compared the construction cost of the test case with that of the retrieved case. The model that was developed in this study calculated the standard error rate and the prediction accuracy. Equation [3] was used to calculate the standard error rate, and equation [4] to calculate the prediction accuracy.

$$f_{SER}(x) = \frac{\left| V_{Test\_Case} - PV_{Retrieved\_Case} \right|}{V_{Test\_Case}} \times 100 \quad [3]$$

$$f_{PA}(x) = 100 - f_{SER}(x) \quad [4]$$

where, f_{SER} is a function of the standard error rate, V_{Test_Case} is the test case value, $PV_{Retrieved\_Case}$ is the prediction value of the retrieved case, f_{PA} is a function of the prediction accuracy.

4.2 Application of GA

In the study conducted by Koo et al. (2008) and Koo (2007), it was shown that the correlation between case similarity and prediction accuracy is not always proportional. It was also shown that the methods of calculating the attribute weight and attribute similarity are critical factors in the calculation of the case similarity. Therefore, in this study, such factors were defined as optimization parameters, and the following optimization process using GA was established:

▪ **Optimization parameter I : Minimum Criteria for scoring Attribute Similarity (MCAS)**

The previous studies applied a specific value recommended by a software program (i.e., the "Esteem" software recommends 10%) (Kim et al. 2004), but in this study, the software "Evolver" was used to conduct a simulation using GA based on the 0-100% range.

▪ **Optimization parameter II : Range of Attribute Weight (RAW)**

In the study conducted by Koo et al. (2008) and Koo (2007), various methodologies were used to deduce the attribute weight that makes the prediction results more accurate, which include ANN, MRA, and FC. It was found that when the sensitivity coefficient deduced from the ANN model was applied as a methodology for discovering the attribute weight, the prediction accuracy was greater than those of FC, MRA (orig.), and MRA (abs.).

Based on the aforementioned results, the optimization process was applied in this study to calculate the attribute weight, where the target was based on the prediction accuracy. The model that was developed in this study could optimize the value of the attribute weight by itself. The software "Evolver" was used to conduct a simulation based on the 0-100% range.

▪ **Constraint : the Number of Prediction Cases (NPC)**

In this study, the minimum criterion was defined based on the number of prediction cases. Although the average prediction accuracy, which is the standard for evaluating the prediction capacity of a model, is high, the predicted accuracy of a certain case would be extremely low. To obtain consistency, the standard deviation of the prediction accuracy must be controlled. This study developed a model with the exception of the cases detected as outliers.

As shown in the shaded part of Fig. 3, a CBR process was integrated with GA. In the study conducted by Koo et al. (2008) and Koo (2007), a similar process was used, where TAW was set to be the optimization parameter, which is different from this study, where RAW was set to be the optimization parameter. And, since it was found in the previous research that MCAS is important in CBR, MCAS was also set to be the optimization parameter to develop the model in this study.

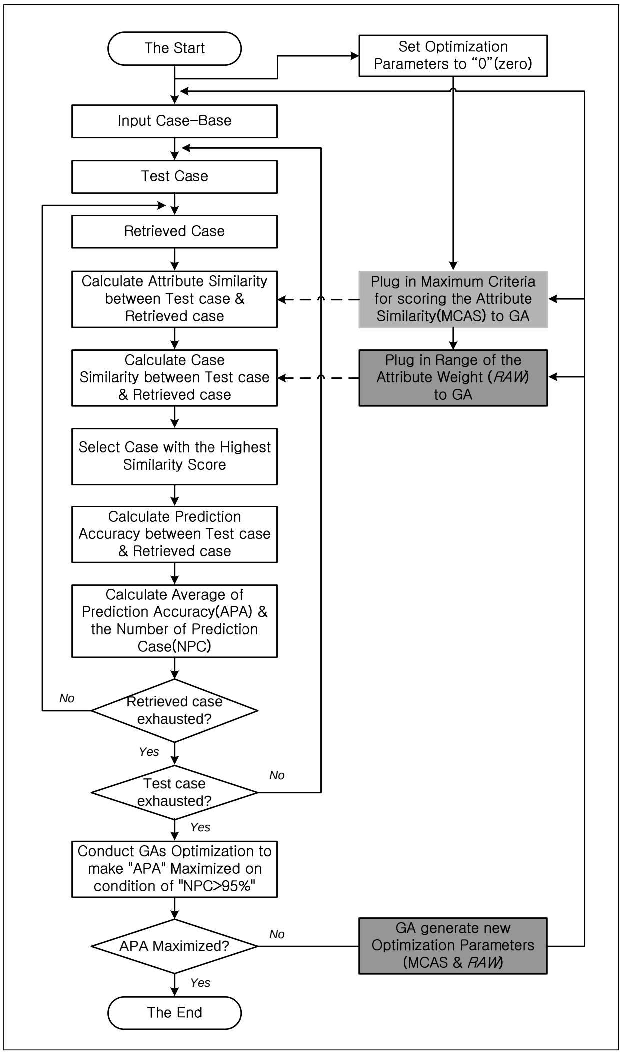

Figure 3. A CBR process integrated with GA

5. RESULTS AND DISCUSSION

5.1 Analysis of MCAS (Optimization Parameter I)

The detailed analysis of the prediction results with regard to the minimum criteria for scoring attribute similarity (MCAS) is as follows (refer to Fig. 4 and 5).

The correlation between MCAS and the prediction accuracy is not always proportional. It was shown that the prediction accuracy goes up and down considerably.

First, as for model I, when the MCAS was set at 77.32%, 87.48%, and 3.79%, respectively, for the structure class, finishing, and others, the prediction accuracy was greatest at 82.649%, 91.409%, and 90.482%, respectively, as shown in Fig. 4.

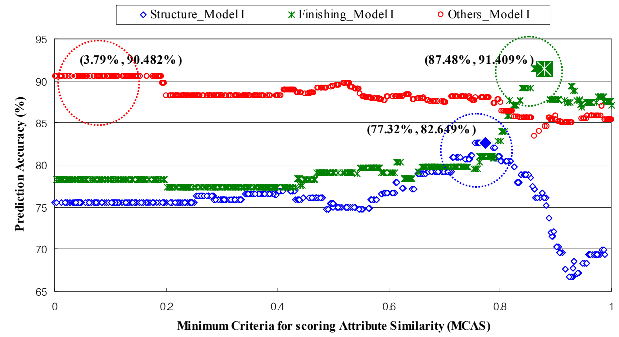

Figure 4. Correlation between MCAS and prediction accuracy in model I

Second, as for model II, when the MCAS was set at 78.58%, 91.93%, and 79.21%, respectively, for the structure class, finishing, and others, the prediction accuracy was greatest at 82.518%, 92.985%, and 91.433%, respectively, as shown in Fig. 5.

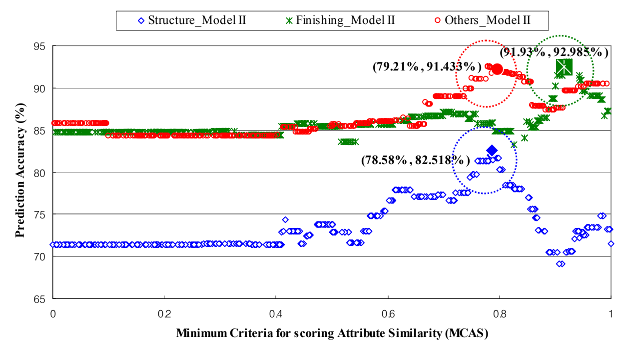

Figure 5. Correlation between MCAS and prediction accuracy in model II

5.2 Analysis of RAW (Optimization Parameter II)

The detailed analysis of the prediction results with regard to the range of attribute weights (RAW) is as follows (refer to Table 4).

The value of the attribute weight by model was derived when the prediction accuracy was greatest. As the database or project information may be changed, the optimization process of the model can be reactivated to find the optimization value.

Table 4. Value of Optimization Parameters by Model

(1) Optimization Parameters		Architecture_Structure				Architecture_Finishing				Others (Landscape, Earth, Mech, Elec, Communication)			
		(2)Model I		(3)Model II		(4)Model I		(5)Model II		(6)Model I		(7)Model II	
		FC	GA	FC	GA	FC	GA	FC	GA	FC	GA	FC	GA
Attribute Weight	A1	1	0.1114	1	0.6670	1	0.4155	1	0.6869	1	0.9683	1	0.0314
	A2	1	0.5730	1	0.2231	1	0.2231	1	0.4759	1	0.2231	1	0.5038
	A3	1	0.0027	1	0.0027	1	0.0027	1	0.4939	1	0.0027	1	0.0027
	A4	1	0.2019	1	0.0113	1	0.0098	1	0.0113	1	0.2214	1	0.0519
	A5	1	0.9320	1	0.8657	1	0.1699	1	0.5310	1	0.1682	1	0.4471
	A6	1	0.2426	1	0.1954	1	0.1101	1	0.0975	1	0.1954	1	0.7721
	A7	1	0.5583	1	0.4164	1	0.4092	1	0.2526	1	0.4164	1	0.4164
	A8	1	0.5142	1	0.4142	1	0.1584	1	0.0014	1	0.2418	1	0.0762
	A9	1	0.0212	1	0.0511	1	0.8726	1	0.4157	1	0.0627	1	0.6920

	A10	1	0.4293	1	0.2828	1	0.5219	1	0.4635	1	0.6690	1	0.0783
	A11	1	0.2940	1	0.3032	1	0.9829	1	0.0739	1	0.9763	1	0.9158
	A12	-	-	1	0.2617	-	-	1	0.0667	-	-	1	0.2617
	A13	-	-	1	0.6268	-	-	1	0.3821	-	-	1	0.5241
	A14	-	-	1	0.6434	-	-	1	0.0017	-	-	1	0.0731
	A15	-	-	-	-	-	-	1	0.0211	-	-	-	-
	A16	-	-	-	-	-	-	1	0.6329	-	-	-	-
	A17	-	-	-	-	-	-	1	0.5826	-	-	-	-
	A18	-	-	-	-	-	-	1	0.2348	-	-	-	-
	A19	-	-	-	-	-	-	1	0.0320	-	-	-	-
	A20	-	-	-	-	-	-	1	0.0011	-	-	-	-
	A21	-	-	-	-	-	-	1	0.6767	-	-	-	-
	A22	-	-	-	-	-	-	1	0.2447	-	-		0.0263
	A23	-	-	-	-	-	-	1	0.5270	-	-		0.2231
	A24	-	-	-	-	-	-	1	0.0152	-	-		0.4183
	A25	-	-	-	-	-	-	1	0.0185	-	-		0.0796
	A26	-	-	-	-	-	-	1	0.1278	-	-		0.2336
	A27	-	-	-	-	-	-	1	0.4737	-	-		0.0995
MCAS		0.7732	0.7732	0.7858	0.7858	0.8748	0.8748	0.9193	0.9193	0.0379	0.0379	0.7921	0.7921
PREDICTION ACCURACY		72.310	82.649	72.287	82.518	85.536	91.409	86.336	92.985	83.495	90.482	83.465	91.433

▪ Model Ⅰ : a model that uses the attributes from A1 to A11
▪ Model Ⅱ : a model that uses the attributes from A1 to A11 and that is selectively applied form A12 to A27 according to the model
▪ A1: Plottage, A2: Total floor area, A3: Land ratio, A4: Floor space index, A5: No. of stories below the ground, A6: No. of stories above the ground, A7: No. of parking lot, A8: Landscape area, A9: Public open space, A10: Facility function, A11: Site Location, A12: Type of Structure(Reinforced concrete), A13: : Type of Structure(Steel & reinforced concrete), A14: Type of Structure(Steel), A15: Type of window(Low-E), A16: Type of window(Universal), A17: Type of glass(Clarity), A18: Type of glass(Color), A19: Type of glass(Reflection), A20: External materials(Metal), A21: External materials(Stone), A22: Grade on environment(I), A23: Grade on environment(II), A24: Grade on environment(None), A25: Grade on communication(I), A26: Grade on communication(II), A27: Grade on communication(None)

In conclusion, a CBR model should be able to optimize the prediction accuracy by itself by finding the optimization value of such parameters as MCAS and RAW using GA. As mentioned earlier, an engine for improving the prediction accuracy of a CBR model was applied to the model in this study. Through future researches, the prediction capability of the proposed cost estimating method could be further improved.

5.3 Analysis of the Prediction Accuracy of the Proposed Cost Model

▪ Average prediction accuracy by CBR model

As shown in Fig. 6, in the case of Architecture_Structure, although the prediction accuracy values of models I and II were not remarkably different, when GA was used to calculate the attribute weight, the prediction accuracy was improved and became higher than that of FC. In the cases of Architecture_Finishing and Others, model II was more predictive than model I, and when GA was used to calculate the attribute weight, the prediction accuracy was improved and became higher than that of FC.

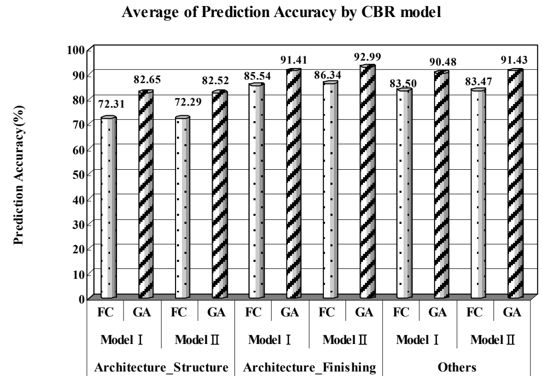

Figure 6. Average of prediction accuracy by CBR model

▪ Standard deviation of prediction accuracy by CBR model

As shown in Fig. 7, in all the cases (Architecture_Structure, Architecture_Finishing, and Others), the standard deviation of model II decreased more than that of model I, and when GA was used to calculate the attribute weight, the standard deviation declined more than that of FC.

It was shown that when some values need to be predicted, the fact that there are more information makes it more accurate and less deviant.

It was also shown that the method to be used for calculating the attribute weight is critical, and that a CBR model should be able to optimize the attribute weight by itself, using GA.

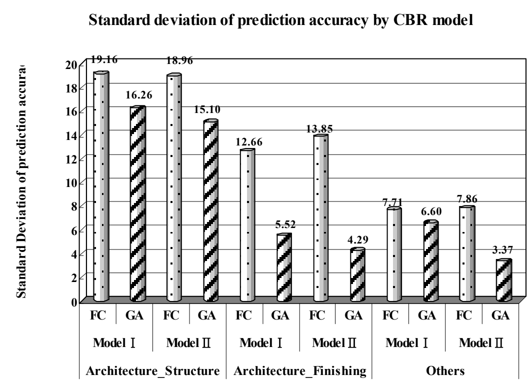

Figure 7. Standard deviation of prediction accuracy by CBR model

Table 5 shows the results of the descriptive analysis with regard to the prediction accuracy by methodology. As shown in the fourth column [(4) Mean] of Table 5, the value of the prediction accuracy in Architecture_Structure was greatest at 82.649% in model I when GA was used to calculate the attribute weight. The value in Architecture_Finishing was greatest at 92.985% in model II when GA was used, and the value of Others was greatest at 91.433% when GA was used.

A slight difference may occur as the number of influencing factors may be changed. The model, however, where GA was used to calculate the attribute weight, was almost more predictive than FC. Moreover, when GA was used to calculate the attribute weight, the standard deviation declined more than that of FC.

It was thus proven that GA could improve the prediction capability (i.e., prediction capacity means both prediction accuracy and standard deviation) of a CBR model.

Table 5. Results of the Descriptive Analysis by CBR Model

(1) Type of Model		(2) Attribute Weight	(3) No. of Cases	(4) Mean	(5) Standard Deviation	(6) Median	(7) Min.	(8) Max.	(9) 5th Percentile
Architecture _Structure	Model I	FC	23	72.310	19.161	74.349	24.901	99.853	48.559
		GA	22	82.649	16.255	87.930	54.205	99.853	55.880
	Model II	FC	23	72.287	18.960	69.848	24.901	99.853	48.558
		GA	23	82.518	15.097	87.376	53.271	99.853	54.785
Architecture _Finishing	Model I	FC	23	85.536	12.664	88.862	54.705	99.944	64.004
		GA	23	91.409	5.516	92.661	76.799	99.944	83.010
	Model II	FC	23	86.336	13.854	90.874	48.851	99.944	55.920
		GA	23	92.985	4.285	93.327	84.094	99.944	85.767
Others	Model I	FC	17	83.495	7.707	83.651	67.203	97.823	73.142
		GA	17	90.482	6.603	91.596	77.291	98.337	78.574
	Model II	FC	17	83.465	7.862	83.895	65.999	94.858	71.929
		GA	17	91.433	3.365	92.132	83.895	98.337	86.671

6. VALIDATION

Table 6, which shows the retrieved case that was the most similar to the test case as to model I, contains not only the predicted value of the construction cost but also the project characteristics of both the test case and the retrieved case. These results may be used as references in the decision-making process. When the case of no. 1 was applied to test case, respectively, for structure, finishing, and others, the retrieved case was the case of no. 2 for all the class. The prediction accuracy was shown at 98.904%.

Table 6. The Case Retrieved by CBR Model I

(1) Optimization Parameters		Structure		Finishing		Others	
		Test Case	Retrieved Case	Test Case	Retrieved Case	Test Case	Retrieved Case
Case No.		1	2	1	2	1	2
Attribute	A1	8908.9	16,604.22	8908.9	16,604.22	8908.9	16,604.22
	A2	3237.99	39,399.12	3237.99	39,399.12	3237.99	39,399.12
	A3	43.49	37.25	43.49	37.25	43.49	37.25
	A4	219.65	137.94	219.65	137.94	219.65	137.94
	A5	2	2	2	2	2	2
	A6	9	12	9	12	9	12
	A7	253	307	253	307	253	307
	A8	1443.31	3,661.30	1443.31	3,661.30	1443.31	3,661.30
	A9	966.54	2,100.00	966.54	2,100.00	966.54	2,100.00
	A10	2	2	2	2	2	2
	A11	1	1	1	1	1	1
CONSTRUCTION COST (₩/㎡)		344,0 94.17	334,9 20.02	437,9 99.39	390,3 58.28	773,3 50.62	813,1 17.18
PREDICTION ACCURACY(%)		97.334		89.123		94.858	
		98.904					

Table 7 shows the retrieved cases that were the most similar to the test case as to model II. The prediction accuracy was shown at 98.904% in the case of no. 1.

Table 7. The Case Retrieved by CBR Model II

(1) Optimization Parameters		Structure		Finishing		Others	
		Test Case	Retrieved Case	Test Case	Retrieved Case	Test Case	Retrieved Case
Case No.		1	2	1	2	1	2
Attribute	A1	8908.9	16,604.22	8908.9	16,604.22	8908.9	16,604.22
	A2	3237.99	39,399.12	3237.99	39,399.12	3237.99	39,399.12
	A3	43.49	37.25	43.49	37.25	43.49	37.25
	A4	219.65	137.94	219.65	137.94	219.65	137.94
	A5	2	2	2	2	2	2
	A6	9	12	9	12	9	12
	A7	253	307	253	307	253	307
	A8	1443.31	3,661.30	1443.31	3,661.30	1443.31	3,661.30
	A9	966.54	2,100.00	966.54	2,100.00	966.54	2,100.00
	A10	2	2	2	2	2	2
	A11	1	1	1	1	1	1
	A12	0	0	0	0	0	0
	A13	1	1	1	1	1	1
	A14	0	0	0	0	0	0
	A15	-	-	1	1	-	-
	A16	-	-	0	0	-	-
	A17	-	-	0	1	-	-
	A18	-	-	1	0	-	-
	A19	-	-	0	0	-	-
	A20	-	-	1	1	-	-
	A21	-	-	1	1	-	-
	A22	-	-	0	0	0	0
	A23	-	-	1	1	1	1
	A24	-	-	0	0	0	0
	A25	-	-	0	0	0	0
	A26	-	-	1	0	1	0
	A27	-	-	0	1	0	1
CONSTRUCTION COST (₩/㎡)		344,09 4.17	334,92 0.02	437,99 9.39	390,35 8.28	773,35 0.62	813,11 7.18
PREDICTION ACCURACY(%)		97.334		89.123		94.858	
		98.904					

7. CONCLUSIONS

In this study, a CBR model integrated with GA was developed based on the characteristics of public-office projects. Especially, to improve the prediction capacity of the CBR model, this study defined the minimum criteria for scoring attribute similarity (MCAS) and the range of attribute weights (RAW) as the optimization parameters, and the optimization process was completed using GA.

As mentioned, it was shown that the prediction accuracy was most accurate when GA was applied as the method of calculating the attribute weight rather than FC. It is expected that the prediction accuracy can be improved through the use of GA in the future (refer to the fourth column in Table 5: “(4) Mean”).

The proposed model is a useful tool for reasonable decision making. It is expected that this model help stakeholders in charge of estimating the budget in a public office at the early stages of a construction project..

To solve the problem of the correlation between case similarity and prediction accuracy not always being proportional, and to make the prediction capacity more accurate, the optimization parameters directly related to the prediction accuracy should be introduced in the following future researches:

- a research related to an engine for filtering the predicted value (i.e., for filtering the predicted value based on the predicted value of either MRA or ANN).
- a research related to the number of cases that should be finally selected to improve the prediction accuracy.

ACKNOWLEDGEMENT

This research was supported by a grant (06 CIT A03) from Construction Infra Technology Program funded by Ministry of Construction & Transportation of Korean government.

REFERENCES

[1] Dogan, S. Z., Arditi, D., and Gunaydin, H. M. (2006). Determining Attribute Weights in a CBR Model for Early Cost Prediction of Structural Systems. *J. Constr. Eng. Manage., 132*(10): 1092-1098.

[2] Duverlie, P., and Castelain, J. M. (1999). Cost Estimation During Design Step : Parametric Method versus Case Based Reasoning Method. *Adv. Manuf. Technol., 15*(12): 895-906.

[3] Hegazy, T., and Ayed, A. (1998). Neural Network Model for Parametric Cost Estimation of Highway Projects. *J. Constr. Eng. Manage., 124*(3): 210-218.

[4] Koo, C. (2007). A CBR-based hybrid model for Predicting Construction Duration and Cost based on Project Characteristics in Multi-Family Housing Projects. M.S. Thesis, University of Seoul, Seoul, Korea.

[5] Koo, C., Hong, T., Hyun, C., and Koo, K. (2008). A CBR-Based Hybrid Model for Predicting a Construction Duration and Cost based on Project Characteristics in Public Multi-Family Housing Projects. Under review in the *Journal of Canadian Journal of Civil Engineering*.

[6] Phaobunjong, K. (2002). Parametric Cost Estimating Model for Conceptual Cost Estimating of Building Construction Projects. PhD thesis, Univ. of Texas, Austin, Tex.