Parameter Impact Applied Case-based Reasoning Cost Estimation

Joseph Ahn[1]; Hyun-Soo Lee[2]; Moonseo Park[3]; Sae-Hyun Ji[4] and Sooyoung Kim[5]

[1] Ph.D. Student, Dept. of Architecture, Seoul National Univ., Seoul, Korea
[2] Prof., Dept. of Architecture, Seoul National Univ., Seoul, Korea
[3] Prof., Dept. of Architecture, Seoul National Univ., Seoul, Korea
[4] Research Assistant, Dept. of Architecture, Seoul National Univ., Seoul, Korea
[5] Ph.D. Student, Dept. of Architecture, Seoul National Univ., Seoul, Korea
Correspond to josephahn@snu.ac.kr

ABSTRACT: To carry out a one-off construction project successfully, effective and accurate early cost estimation is crucial, especially during the conceptual stage where very limited minimum information of construction project is given. As the level of accuracy of the early cost estimation has huge impacts on precise budgeting and cost management of a project, in other words, reducing the risk of a project, cost must be managed with special awareness. In an effort to improve the estimate accuracy of cost during the conceptual stage, this research introduces a Parameter Impact (PI) which can quantify weights of parameters and rank them; and PI development derived from the principle of impulse in physics is explicated. For a case study, 76 public apartment building cases in Korea are analyzed. To examine the validity of the proposed PI, a validation in terms of CBR applicability test and estimate accuracy comparisons using 10-nearest neighbor cases are carried out. The validation results support that the suggested PI can be applied in quantifying the weights of the parameters and CBR method for early cost estimation.

Keywords: Cost Estimation; Case-based Reasoning; Parameter, Weight

1. INTRODUCTION

Construction project has some unique characteristics in terms of the building type, scope, quality, cost, duration, and so on. All of these necessities of a client must be achieved satisfactorily by contractors and consultants. However, it is notable that there is increasing demands to higher quality for shortened duration at lower cost. In general, a substantial amount of cost is incurred to accomplish a one-off construction project successfully. As insisted by Kim (2005), entire or partial construction cost should be regarded as one of key decision making elements and decided in the early design stages. Even though cost estimation and cost checking activities are conducted during the entire construction phases, more accurate cost estimate is required by owners in the early design stages due to its huge impacts on cost reduction (Kim 2005).

As an effort to improve the estimate accuracy during the conceptual and schematic design stages, various estimation methods have been constantly researched and developed. Among them, the most commonly used method is the parametric estimation which utilizes parameters developed from the historical cost database and the construction practitioners. Apart from the parametric method, the case-based reasoning (CBR) which utilizes the knowledge and experience obtained from the past have been introduced to the construction industry and applied as an estimation method of the construction cost (An et al. 2007, Chou 2009, Ji et al 2011a).

When CBR method is used to estimate the cost, extracting influential parameters on cost and measuring them in a quantitative manner is vital issue. However, there are numerous parameters affecting cost thus if these are all used to estimate, not only the accuracy of the early cost estimation decreases, but also the efficiency of the estimate process becomes more complicated. Therefore, it is necessary to measure and prioritize impacts of parameters in the order of high cost implications. As means of prioritizing variables, various methods such as analytic hierarchy process (AHP), genetic algorithm (GA), principal component analysis (PCA), feature counting (FC), and correlation analysis (CA) are commonly used. Using these methods, weights of parameters are calculated.

However, through literature reviews it is found that calculation process of getting weights by these methods is carried out without considering ranges that each parameter can have. CBR based cost estimation models are generally made based on attained historical database.

As an effort to deal with this challenging issue, this research introduces a Parameter Impact (PI) which can measure weights of parameters quantitatively. After conducting comprehensive literature reviews, a necessity of this research is being explained; and then the PI which adopts the principle of impulse in physics is developed in following section. For a case study, 76 public apartment building cases in Korea are collected and analyzed. To validate the logic of the suggested PI, a validation on CBR applicability issue and its accuracy is carried out.

2. Preliminary Research

CBR method which came from the cognitive science is increasingly being applied to cost estimating. This method utilizes experience of past cases to work out new ones (Schirmer 2000) and is capable of supporting decisions (Pal and Shiu 2004). It also helps improve the accuracy of the cost estimation of building projects and provides solutions for estimating construction cost using numerous parameters which have non-linear relationship with cost (Chou 2009). However, computing similarity required during the retrieval process still remains as a challenging issue (Pal and Shiu 2004). Also, assigning the attribute weight values necessary to search the most similar case needs to be examined (Ji et al 2011a).

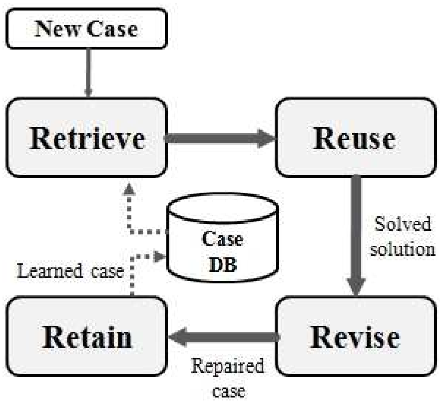

Figure 1. CBR Cycle(Aamodt & Plaza, 1994)

An et al. (2007) proposed a case-based reasoning cost estimating model and attempted to include experience using analytic hierarchy process. A questionnaire survey was conducted and AHP approach was utilized to elicit the domain knowledge from experts and determine the weights of attributes. Dogan et al. (2006) introduced a spreadsheet based CBR prediction model of structural systems and assessed the performance of it by testing the impact of attribute weights generated by three different techniques, namely feature counting, gradient descent, and genetic algorithm.

As reviewed, when cost estimates are conducted using the CBR method, how many and which parameters will be used for estimation are very critical and these are still on-going research topics. Some researches adopted the intuitive approach which relies on the experience of construction industry practitioners. However, as this method is based on subjective aspects of experience and knowledge, a selecting process of dominant parameters might not be very reliable compared to other methods capable of quantifying parameters (Duverlie and Castelain 1999). Some utilized correlation coefficients, genetic algorithms, feature counting, regression coefficients significance or analytic hierarchy process to determine parameter weights and extract influential parameters. However, it is important to note that cost estimating models are built based on obtained historical cost database. According to the size and accuracy of the cost database, parameters weights, cost estimating equations and cost models can be affected to a certain level. Therefore, this research arise a necessity of a parameter weight assigning method which can consider the relationships between parameters and the size of the database apart from mere relationships between parameters and costs.

3. Parameter Impact Development

To deal with this issue, this research introduces a Parameter Impact (PI) which can help measuring weights of parameters and prioritizing them. A PI can be attained by a weight of a parameter multiplied by its range. This is based on the assumption that a range of each parameter is different from others and has different cost implication. Such an assumption is derived from the principle of impulse in physics, which is equal to momentum multiplied by the time for the force is applied (Eq. 1). Impulse is proportional to momentum and momentum is in turn proportional to mass and velocity (Eq. 2). Momentum cannot be obtained only by the mass of an object and varies depending on the velocity as well. As velocity is also in proportion to displacement of an object when time is fixed (Eq. 3), it is found that momentum is proportional to mass and displacement. Adopting the aforementioned principle, mass of an object is being substituted by weight of parameter and displacement by range. To sum up, it is defined that Parameter Impact (PI) is determined by a weight of parameter times a range of parameter (Eq. 4).

$$\mathbf{I=F \cdot \Delta t} \qquad \text{(Eq.1)}$$

I: impulse F: momentum

$$\mathbf{F = m \cdot V} \qquad \text{(Eq. 2)}$$

F: momentum m: mass V: velocity

$$\mathbf{Vav} = \frac{\Delta r}{\Delta t} \qquad \text{(Eq. 3)}$$

Δr: displacement of an object Δt: a time interval

$$\textbf{Parameter Impact } (\mathbf{PI_i}) = |\, \mathbf{PW_i * PR_i} \,| \qquad \text{(Eq. 4)}$$

4. Data Acquisition and Analysis

Under housing supply legislation, apartment households in Korea have been produced by unit gross area. Thus, cost data of apartment building projects can be analyzed and the database can be built according to unit types as similar patterns are expected to be obtained. The data analysis of the research is conducted based on the building cost data (refer Table 1), called priced bills of quantities, of 76 apartment building cases in Korea.

These data are provided by Seoul Housing Corporation, Korean public enterprise established in 1989. To elaborate, priced bills of quantities which contain the total expenditure including all inputs such as labor and materials are prepared by public owners and utilized as a vital standard to estimate a precise budget. These priced bills of quantities are also used by construction firms to determine a total-fixed-price for their bid proposals and awards are made based on it.

Table 1. Profile of Cost Data

Project	A	B	C	total
year	2005	2007	2008	
Number of cases (Type 84)	50	9	17	76
Number of households	2076	190	222	2,488

Through data analysis, it is found that there are four types of apartment households (type $49m^2$, $59m^2$, $84m^2$ and $114m^2$) and apartment buildings are comprised of either singular types or a combination of different types. To improve the convenience of data analysis and an accuracy of estimation of building cost, all the historical data are separated and classified into the singular type of unit gross area according to the each unit type. Furthermore, the gross area ratio is adopted into takeoff priced quantities of each item as priced bills of quantities are structured with trade sections and grouped by the sum of quantities of each input item.

5. Computing Process

A pool of more than a hundred parameters was created from the drawings available in the early design stages. Among the pool, eleven parameters (number of households, gross floor area, number of unit floor households, number of elevators, number of floors, number of piloti with household scale, number of households of unit floor per elevator, height between stories, depth of pit, roof type, hallway type), which are considered to have significant impacts on cost, are chosen by construction practitioners who had participated in the construction industry more than fifteen years.

To calculate values of PI, this research adopts correlation coefficients as weights of the parameters. The quantification of a correlation is usually executed by specifying the correlation coefficient. In statistics, a correlation refers to the departure of two variables from independence, while the correlation coefficient indicates the strength and direction of a linear relationship between two random variables. Generally, a correlation coefficient of under -0.5 or over 0.5 indicates that the two variables have a strong correlation.

Regarding a range of a parameter, it is obtained by maximum value minus minimum value of each parameter from the database. Consequently, the Parameter Impact (PIi) is derived by multiplying a correlation coefficient (RCi) by a range of data of a parameter (PRi), as below (Eq. 5). Derived values of PIs, correlation coefficients and ranges are summarized; and their rankings are described in Table 2.

Parameter Impact (PI_i) = | CC_i * PR_i | (Eq. 5)

Table 2. Derived Value of PI & CC and their Rankings

Parameter	CC weight	CC Rank	Range	PI weight	PI Rank
X1	0.97	2	51	49.57	2
X2	0.98	1	5,633.99	5,521.55	1
X3	0.9	3	3	2.71	5
X4	0.41	10	1.5	0.62	9
X5	0.79	4	11	8.67	3
X6	0.56	9	6	3.36	4
X7	0.7	6	2	1.4	6
X8	0.71	5	0.1	0.07	11
X9	0.03	11	4.36	0.13	10
X10	0.62	8	1	0.62	8
X11	0.68	7	1	0.68	7

Note: CC: Correlation Coefficient, Range = Maximum value – Minimum value, PI: Parameter Impact, (X1) Number of households, (X2) Gross floor area, (X3) Number of unit floor households, (X4) Number of elevators, (X5) Number of floors, (X6) Number of piloti with household scale, (X7) Number of households of unit floor per elevator, (X8) Height between stories, (X9) Depth of pit, (X10) Roof type, (X11) Hallway type

6. Validation

A validation is a PI applicability test to CBR estimation method. As illustrated in Figure 1, to measure and score the similarity of the retrieved cases in CBR based cost estimation, weights of parameters and a similarity function are required. It is critical to have accurate and reliable parameter weights and a similarity function as the accuracy of similarity measurement is highly dependent on them.

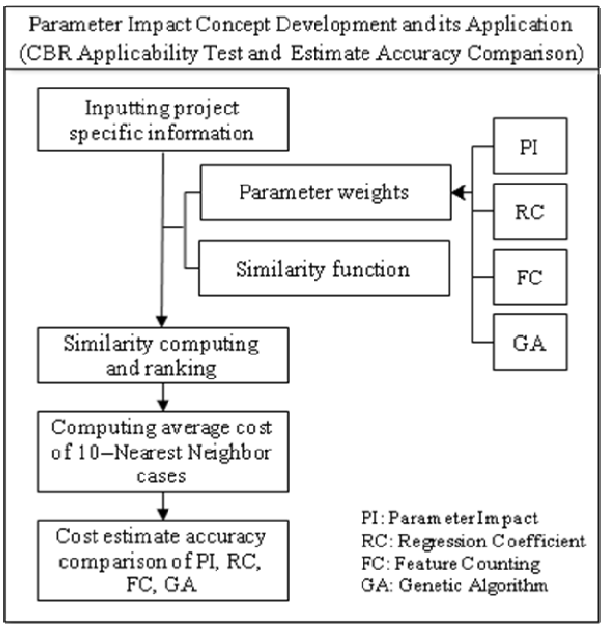

Figure 2. Experiment Design for Validation

After having closer examinations concerning the similarity function, it is found that Euclidean distance, the square root of the sum of the square of the arithmetical differences between two corresponding objects, is the most commonly used distance measuring method (Pal and Shiu 2004). Hence, this research adopts the similarity measuring function based on Euclidean distance concept employed by Ji et al. (2011b).

Apart from the similarity function mentioned above, parameter weights are also required for similarity measurement. This validation is especially designed to examine the validity of the suggested PI as a means of assigning weights of parameters. CBR based cost estimation using weights of PI is compared to that of standardized regression coefficient (RC), feature counting (FC) and genetic algorithm (GA). Consequently, absolute error ratios (AER) of 10-NN using each parameter weight methods are compared (refer Table 4).

Table 3. Profile of Test Cases

Type 84	Parameter						Total cost
	X1 X7	X2 X8	X3 X9	X4 X10	X5 X11	X6	
A	13.0 2.0	1,416 2.9	2.0 7.6	1.0 -	7.0 1.0	1.0	970,004,637
B	44.0 2.0	3,368 2.8	4.0 5.0	2.0 -	12.0 1.0	4.0	2,094,424,957
C	44.0 2.0	4,890 2.8	4.0 9.4	2.0 -	12.0 1.0	2.0	3,038,402,053
D	50.0 4.0	5,500 2.8	4.0 5.9	1.0 1.0	13.0 -	2.0	3,220,935,698
E	56.0 4.0	6,168 2.8	4.0 5.1	1.0 1.0	15.0 -	4.0	3,731,645,455

Note: (X1) Number of households, (X2) Gross floor area, (X3) Number of unit floor households, (X4) Number of elevators, (X5) Number of floors, (X6) Number of piloti with household scale, (X7) Number of households of unit floor per elevator, (X8) Height between stories, (X9) Depth of pit, (X10) Roof type, (X11) Hallway type

Table 4. CBR Applicability Test

Case	PI	RC	FC	GA
	10-NN	10-NN	10-NN	10-NN
A	7.09%	10.15%	5.52%	19.27%
B	9.95%	6.71%	8.44%	15.51%
C	1.13%	8.10%	1.93%	9.35%
D	5.59%	1.58%	24.54%	2.19%
E	7.94%	7.06%	7.43%	6.46%
Average	**6.34%**	**6.72%**	**9.57%**	**10.56%**
S.D	**3.31%**	**3.17%**	**8.73%**	**6.87%**

Note: PI: Parameter Impact, RC: Standardized Regression Coefficient, FC: Feature Counting, GA: Genetic Algorithm

7. CONCLUSIONS

To support decision-makings of a client, cost estimation is conducted and higher estimate accuracy, especially in the early design stages, is getting required to maximize stability of cost planning. In an effort to improve estimate accuracy during the initial stages, this research introduced the Parameter Impact (PI) which can measure the weights of parameters in quantitative manner and prioritize them so that the rankings in terms of cost implication can be determined. To validate the logic of PI, the validation tested the reliability of PI as a method of assigning weights of parameters by comparing estimate accuracy with that other methods such as standardized regression coefficient, feature counting and genetic algorithm when CBR based cost estimation is utilized.

The validation results support that the suggested Parameter Impact can be utilized in measuring weights of parameters in quantitative manner; and it can leads to improvement of estimate accuracy when the CBR estimation is performed. However, this research was limited to public apartment buildings in Korea; thus the PI should be validated with other building types for generalization. More importantly, the PI concept itself needs to be further examined and developed.

ACKNOWLEDGEMENT

This research was supported by grants (R&D06CIT-A03) from the Innovative Construction Cost Engineering Research Center and (code # 09 R&D A01) from Super-Tall Building R&D Project funded by the Ministry of Land, Transport and Marine Affairs of Korean Government.

REFERENCES

[1] An, S-H. et al., "A Case-based Reasoning Cost-estimating Model Using Experience by Analytic Hierarchy Process," *Building and Environment,* Vol. 42*,* pp. 2573–2579, 2007.

[2] Chou, J-S., "Web-based CBR System Applied to Early Cost Budgeting for Pavement Maintenance Project," *Expert System with Applications,* Vol. 39*,* pp. 2947-2960, 2009.

[3] Dogan, S. Z. et al., "Determining Attribute Weights in a CBR Model for Early Cost Prediction of Structural System," *Journal of Construction Engineering and Management,* Vol. 132(10), ASCE, pp. 1092–1098, 2006.

[4] Duverlie, P. and Castelain J. M., "Cost Estimation during Design Step: Parametric Method versus Case-based Reasoning Method," *International Journal of Advanced Manufacturing Technology,* Vol. 15*,* pp. 895–906, 1999.

[5] Ji, S. H. et al.(a), "Military Facility Cost Estimation System (MilFaCE) Using Case-Based Reasoning in Korea," *Journal of Computing in Civil Engineering*, Vol. 25(3), pp. 218-231, 2011.

[6] Ji, S. H. et al.(b), "Cost Estimation Model for Building Projects Using Case-Based Reasoning," *Canadian Journal of Civil Engineering,* Vol. 38, pp. 570-581, 2011.

[7] Kim, M. H., *Cost Planning in Architecture*, Kimoondang, Seoul, 2005.

[8] Pal, S. K. and Shiu, S. C. K., *Foundations of Soft Case-Based Reasoning*, Wiley Interscience, 2004.

[9] Schirmer, A., "Case-Based Reasoning and Improved Adaptive Search for Project Scheduling," *Naval Research Logistics,* Vol. 47(3), pp. 201–222, 2000.